

FD - 610111, 121 14291000
PD - 12/01/99
sk 60168229

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 18, 2003, 05:32:01 ; Search time 3098 Seconds
(without alignments)
10342.863 Million cell updates/sec

Title: US-09-874-062-2
Perfect score: 1101
Sequence: 1 ctgggaaccaccacccagacc.....caaaaaaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hgt_hum:*
- 31: em_hgt_inv:*
- 32: em_hgt_other:*
- 33: em_hgt_mus:*
- 34: em_hgt_pln:*
- 35: em_hgt_rod:*
- 36: em_hgt_vrt:*
- 37: em_hgt_vrt:*
- 38: em_sy:*
- 39: em_hgtgo_hum:*
- 40: em_hgtgo_mus:*
- 41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	909.2	82.6	347572	2	AC107303	AC107303 Homo sapi
2	876.8	79.6	161987	9	AC006070	AC006070 Homo sapi
3	735.6	66.8	174032	2	AC025904	AC025904 Homo sapi
4	669	60.8	671	9	HS406935	AJ406935 Homo sapi
5	403.4	36.6	133392	2	AC099183	AC099183 Rattus no
6	403.4	36.6	182806	2	AC098902	AC098902 Rattus no
7	378.2	34.4	192027	2	AC067715	AC067715 Mus muscu
8	378.2	34.4	200602	10	AL590997	AL590997 Mouse DNA
9	368.2	33.4	197658	2	AL646094	AL646094 Mus muscu
10	368.2	33.4	200602	10	AL590997	AL590997 Mouse DNA
11	356	32.3	174032	2	AC025904	AC025904 Homo sapi
12	356	32.3	347572	2	AC107303	AC107303 Homo sapi
13	354.4	32.2	1099	9	HS406936	AJ406936 Homo sapi
14	354.4	32.2	1203	9	HS406939	AJ406939 Homo sapi
15	354.4	32.2	186078	9	AC037482	AC037482 Homo sapi
16	343.8	31.2	197658	2	AL646094	AL646094 Mus muscu
17	335.4	30.5	1220	9	HS406944	AJ406944 Homo sapi
18	335.4	30.5	186078	9	AC037482	AC037482 Homo sapi
19	335.2	30.4	111637	2	AC100808	AC100808 Homo sapi
20	328.4	29.8	1093	9	HS406943	AJ406943 Homo sapi
21	324.2	29.4	111637	2	AC100808	AC100808 Homo sapi
22	319	29.0	1080	9	BC004180	BC004180 Homo sapi
23	319	29.0	1080	9	BC004212	BC004212 Homo sapi
24	314	28.5	55568	2	AC015850_3	AC015850_3 Homo sapi
25	312.6	28.4	182806	2	AC098902	AC098902 Rattus no
26	310.4	28.2	118444	10	AL590992	AL590992 Mouse DNA
27	303.2	27.5	924	9	HS4296168	AJ296168 Homo sapi
28	297.4	27.0	942	9	HS406937	AJ406937 Homo sapi
29	297	27.0	110000	2	AC015850_0	AC015850_0 Homo sapi
30	287	26.1	110000	2	AC015850_0	AC015850_0 Homo sapi
31	270.8	24.6	1064	9	HS406941	AJ406941 Homo sapi
32	269.2	24.5	1037	9	HS406945	AJ406945 Homo sapi
33	265.6	24.1	485	4	OAHKP	X73462 O.aries mRN
34	263.2	23.9	133392	2	AC099183	AC099183 Rattus no
35	252	22.9	3489	6	AR072952	AR072952 Sequence
36	252	22.9	3489	6	AX107938	AX107938 Sequence
37	252	22.9	3489	14	KSU52064	U52064 Kaposi's sa
38	252	22.9	32207	6	AR065852	AR065852 Sequence
39	252	22.9	32207	6	AR127850	AR127850 Sequence
40	252	22.9	32207	6	AR194752	AR194752 Sequence
41	252	22.9	137508	14	KSU75698	U75698 Kaposi's sa
42	251.4	22.8	110000	2	AC015850_2	Continuation (3 of
43	250.8	22.8	782	9	HS406934	AJ406934 Homo sapi
44	241.2	21.9	63280	2	AC105960	AC105960 Mus muscu
45	238.2	21.6	28559	14	AF148805	AF148805 Kaposi's

ALIGNMENTS

RESULT 1
AC107303
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-595122, WORKING DRAFT
ACCESSION AC107303
VERSION AC107303.3 GI:20335965
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 347572)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

AC107303 347572 bp DNA linear HTG 09-MAY-2002
Homo sapiens chromosome 3 clone RP11-595122, WORKING DRAFT
SEQUENCE, 32 unordered pieces.

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Galis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Haves, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczka, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 347572)
 Worley, K.C.

Direct Submission
 Submitted (18-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 347572)
 Worley, K.C.

Direct Submission
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 28, 2002 this sequence version replaced gi:18449841.

----- Genome Center.
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HDSS
 Center clone name: RP11-595122
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 331666 bases at least Q40
 Consensus quality: 339305 bases at least Q30
 Consensus quality: 344140 bases at least Q20
 Estimated insert size: 172965; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation.

TITLE JOURNAL REFERENCE JOURNAL AUTHORS JOURNAL

REFERENCE JOURNAL AUTHORS JOURNAL

COMMENT

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 consists of 32 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 2351: contig of 2351 bp in length
 2352 2451: gap of unknown length
 2452 6192: contig of 3741 bp in length
 6193 6292: gap of unknown length
 6293 10339: contig of 4047 bp in length
 10340 10439: gap of unknown length
 13681: contig of 3242 bp in length
 13682 13781: gap of unknown length
 13782 18293: contig of 4512 bp in length
 18294 18393: gap of unknown length
 18394 22934: contig of 4541 bp in length
 22935 23034: gap of unknown length
 26005: contig of 3571 bp in length
 26006 26705: gap of unknown length
 26706 31631: contig of 4925 bp in length
 31632 31730: gap of unknown length
 31731 36997: contig of 5267 bp in length
 36998 37097: gap of unknown length
 37098 41631: contig of 4534 bp in length
 41632 41731: gap of unknown length
 41732 48452: contig of 6721 bp in length
 48453 48552: gap of unknown length
 48553 55906: contig of 7254 bp in length
 55907 55906: gap of unknown length
 55907 65188: contig of 9282 bp in length
 65189 65288: gap of unknown length
 65289 71857: contig of 6569 bp in length
 71858 71957: gap of unknown length
 71958 80512: contig of 8555 bp in length
 80513 80612: gap of unknown length
 80613 87649: contig of 7037 bp in length
 87650 87749: gap of unknown length
 87750 94122: contig of 6373 bp in length
 94123 94222: gap of unknown length
 94223 103064: contig of 8842 bp in length
 103065 103164: gap of unknown length
 103165 112059: contig of 8895 bp in length
 112060 112159: gap of unknown length
 112160 120884: contig of 8725 bp in length
 120885 120984: gap of unknown length
 120985 134158: contig of 13174 bp in length
 134159 134258: gap of unknown length
 134259 146831: contig of 12573 bp in length
 146832 146931: gap of unknown length
 146932 158213: contig of 11282 bp in length
 158214 158313: gap of unknown length
 158314 172201: contig of 13888 bp in length
 172202 172301: gap of unknown length
 172302 184228: contig of 11927 bp in length
 184229 184328: gap of unknown length
 184329 195498: contig of 11170 bp in length
 195499 195598: gap of unknown length
 195599 210746: contig of 15148 bp in length
 210747 210846: gap of unknown length
 210847 226328: contig of 15482 bp in length
 226329 226429: gap of unknown length
 226430 246181: contig of 19753 bp in length
 246182 246281: gap of unknown length
 246282 265510: contig of 19229 bp in length
 265511 265611: gap of unknown length
 265612 293686: contig of 28076 bp in length
 293687 347572: gap of unknown length
 347573 347572: contig of 53786 bp in length.

Location/Qualifiers
 1. 347572

FEATURES
 source

SEQ ID NO: 2, Result 2
Database: GenBank; AC NO: AC006070

/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="3"	
/clone="RP11-595I22"	
BASE COUNT	107326 a 64859 c 63333 g 108900 t 3154 others
ORIGIN	
Query Match 82.6%; Score 909.2; DB 2; Length 347572;	
Best Local Similarity 91.9%; Pred. No. 6e-148;	
Matches 1002; Conservative 0; Mismatches 13; Indels 75; Gaps 1;	
QY	1 CTGGGAACCCAGCCAGCACTCCACCTCTGAGCCATGGTCAGCTCTGCTGGCTCT 60
Db	105475 CTGGGAACCCAGCCAGCACTCCACCTCTGAGCCATGGTCAGCTCTGCTGGCTCT 105534
QY	61 GTCTGCTGTGACGAGCTGTGGTCAAGCTCTCGCCAGCAGAGCTGTGCGGCCCCAGC 120
Db	105535 GTCTGCTGTGACGAGCTGTGGTCAAGCTCTCGCCAGCAGAGCTGTGCGGCCCCAGC 105594
QY	121 TGTGCGCAGACCACTGTGCGAGGACCACTGCTGCGGCCAGCTGTGCTGCAATTTCCAGT 180
Db	105595 TGTGCGCAGACCACTGTGCGAGGACCACTGCTGCGGCCAGCTGTGCTGCAATTTCCAGT 105654
QY	181 TGTGCGAGCCCTTCTGCTGTATCTCCAGCTGCTGCAAAACCACTGCTGACCAACC 240
Db	105655 CCGTCGAGCCCTTCTGCTGTATCTCCAGCTGCTGCAAAACCACTGCTGACCAACC 105714
QY	241 TGTGCGAGCAGCACTGCTGCGGCCAGCTGCTGCAATTTCCAGTGTGCGAGGCTTCC 300
Db	105715 TGTGCGAGCAGCACTGCTGCGGCCAGCTGCTGCAATTTCCAGTGTGCGAGGCTTCC 105774
QY	301 TGTGCTATCTCCAGCTGTGCGAAACCACTGCTGCGAGCAGCACTGCTGCGGCCCCAGC 360
Db	105775 TGTGCTATCTCCAGCTGTGCGAAACCACTGCTGCGAGCAGCACTGCTGCGGCCCCAGC 105818
QY	361 TGTGCTATCTCCAGCTGTGCGAGGCTTCTGCTGTATCTCCAGCTGTGCGAAACCACTG 420
Db	105819 TGTGCTATCTCCAGCTGTGCGAGGCTTCTGCTGTATCTCCAGCTGTGCGAAACCACTG 105819
QY	421 TGTGCGAGCAGCACTGCTGCGGCCAGCTGCTGCAATTTCTAGTGTGCGAGGCTTCC 480
Db	105820 TGTGCGAGCAGCACTGCTGCGGCCAGCTGCTGCAATTTCTAGTGTGCGAGGCTTCC 105879
QY	481 TGTGCTATCTAGCTGTGCGAAACCACTGCTGCGCAGCAGCACTGCTGCGGCCCCAGC 540
Db	105880 TGTGCTATCTAGCTGTGCGAAACCACTGCTGCGCAGCAGCACTGCTGCGGCCCCAGC 105939
QY	541 TGTGCTATCTCCAGCTGTGCGAGGCCAGCTGCTGCGAGGCTGCTGCGGCCCGGCT 600
Db	105940 TGTGCTATCTCCAGCTGTGCGAGGCCAGCTGCTGCGAGGCTGCTGCGGCCCGGCT 105999
QY	601 TGTGCTATCTAGTGTGCTGCTATCCAGCTGCTGTGTCAGCTGCGCGCTGCTTTC 660
Db	106000 TGTGCTATCTAGTGTGCTGCTATCCAGCTGCTGTGTCAGCTGCGCGCTGCTTTC 106059
QY	661 AGTGTGCGCCAGCACTGCTGTAGAACCACTGCTTCCACCCATCTGCTGCGGCACTTCT 720
Db	106060 AGTGTGCGCCAGCACTGCTGTAGAACCACTGCTTCCACCCATCTGCTGCGGCACTTCT 106119
QY	721 TGTGCTGAGTGAAGCTGCTGCTGATTTGTGCACTTCTTGTCTCAACCTTCAGTTCAG 780
Db	106120 TGTGCTGAGTGAAGCTGCTGCTGATTTGTGCACTTCTTGTCTCAACCTTCAGTTCAG 106179
QY	781 GCACAGATGATCTATTACAGAACATGTGGACTTCTGTATGCTGTGAAACAGAGCGATG 840
Db	106180 GCACAGATGATCTATTACAGAACATGTGGACTTCTGTATGCTGTGAAACAGAGCGATG 106239
QY	841 GACTGATTTGGAAATATTTTATGATGATCTATCTCTTTATAGAGTTTATTATTCCTA 900
Db	106240 GACTGATTTGGAAATATTTTATGATGATCTCTCTTTATAGAGTTTATTATTCCTA 106299
QY	901 TTGAATCTGAATTTACAGTCAAAATTCACATGATGATTTTATGATTTCTTTATTTAAAT 960
Db	TTGAATCTGAATTTACAGTCAAAATTCACATGATGATTTTATGATTTCTTTATTTAAAT

Db	106300	TTGAATCTGAATTTTACAGTCAAAATTCACATGATGATTTTAGAACTCTTTTATTTCTAATTT	106359
QY	961	CAATATACATAAATCTTCAAAATGGTATCTTCTAGATGTTTCTTCTCTAATGTTTCTTCTG	1020
Db	106360	CAATATACATAAATCTTCAAAATGGTATCTTCTAGATGTTTCTTCTCTAATGTTTCTTCTG	106419
QY	1021	GTATCAATTTTTCATGTGGAATTTGTGATGTTTCCCAATAAAACCTTCATGTTTCAAAA	1080
Db	106420	GTATCAATTTTTCATGTGGAATTTGTGATGTTTCCCAATAAAACCTTCATGTTTCAAAA	106479
QY	1081	GCAAAAAAAA 1090	
Db	106480	GCAACAAAAA 106489	

RESULT 2	161987 bp	DNA	linear	PK1 17-DBC-1998
AC006070	Homo sapiens chromosome 17, clone hRPK.206_C_20, complete sequence.			
AC006070	AC006070.1	GI:4028942		
DEFINITION	HTG.			
ACCESSION	Homo sapiens.			
VERSION	Homo sapiens			
KEYWORDS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
SOURCE	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
ORGANISM	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,			
REFERENCE	1 (bases 1 to 161987)			
AUTHORS	2 (bases 1 to 161987)			
TITLE	Unpublished			
JOURNAL	Homo sapiens chromosome 17, clone hRPK.206_C_20			
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,			
REFERENCE	2 (bases 1 to 161987)			
AUTHORS	Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,			
TITLE	Brown, A., Castle, A., Cerny, J., Collangelo, M., Collins, S.,			
JOURNAL	Collamore, A., Cooke, P., Corliss, D., Depayre, E., Devon, K., Dewar, K.,			
AUTHORS	Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, K.,			
REFERENCE	Gage, D., Gardyna, S., Geraghty, K., Grant, G., Hagos, B., Heald, A.,			
AUTHORS	Herena, L., Horton, L., Howland, J. C., Jacotot, L., Jones, C., Kamm, L.,			
TITLE	Karatas, A., Lehoucq, J., Macdonald, P., Marquis, N., McEwan, P.,			
JOURNAL	McGurk, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J.,			
AUTHORS	Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T.,			
REFERENCE	O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A.,			
AUTHORS	Severy, P., Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C.,			
TITLE	Subramanian, A., Testaye, S., Tichovolsky, N., Tortorella-Miller, I.,			
JOURNAL	Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D.,			
AUTHORS	Ye, W. J., Zhao, J. and Zody, M.			

Direct Submission	
Submitted (26-NOV-1998)	Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA	
3 (bases 1 to 161987)	
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,	
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,	
Castle, A., Cerny, J., Collangelo, M., Collins, S., Collamore, A.,	
Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K.,	
Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, K.,	
Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heald, A.,	
Herena, L., Horton, L., Howland, J. C., Jones, C., Kamm, L., Karatas, A.,	
Lehoucq, J., Macdonald, P., Marquis, N., McEwan, P., McGurk, A.,	
McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J.,	
Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P.,	
Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A.,	
Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C.,	
Subramanian, A., Testaye, S., Tortorella-Miller, I., Vassiliev, H.,	
Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.	
Direct Submission	
Submitted (17-DEC-1998)	Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA	
On Dec 17, 1998 this sequence version replaced gi:4006846.	
All repeats were identified using RepeatMasker: Smit, A.F.A. &	
Green, P. (1996-1997)	
http://ftp.genome.washington.edu/RN/RepeatMasker.html	

Only the last 161987 base pairs of this clone are being submitted.

Best Local Similarity 94.6%; Pred. No. 2.4e-142; Matches 908; Conservative 0; Mismatches 52; Indels 0; Gaps 0;			
QY	131	CCACCTGCTGCAGGACCACTGCTGCGCCCGCCAGCTGCTGCATTTCCAGTTCCTGCAGCG 190	
Db	138927	CCTCTGACCCCATGCTGACGCTCCTGCTGTGGCTCTGCTGCTCTGACACAGAGCTGTGGTC 138986	
QY	191	CTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCACTGCTGCAGCA 250	
Db	138987	AAGGTCTCGCCAGGAGAGCTGCTGCGCCCGCCAGCTGCTGCCAGACCACTGCTGCAGCA 139046	
QY	251	CCACCTGCTGCGCCCGCCAGCTGCTGCAATTTCCAGTTGCTGCGAGGCTTCCCTGCTGTATCT 310	
Db	139047	CCACCTGCTGCGCCCGCCAGCTGCTGCAATTTCCAGTTGCTGCGAGGCTTCCCTGCTGTATCT 139106	
QY	311	CCAGCTGCTGCAAAACCCAGCTGCTGAGGACCACTGCTGCGCCCGCCAGCTGCTGCATTT 370	
Db	139107	CCAGCTGCTGCAAAACCCAGCTGCTGAGGACCACTGCTGCGCCCGCCAGCTGCTGCATTT 139166	
QY	371	CCAGTTGCTGCGAGGCTTCCCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGCA 430	
Db	139167	CCAGTTGCTGCGAGGCTTCCCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGCA 139226	
QY	431	CCACCTGCTGCGCCCGCCAGCTGCTGCAATTTCTAGTTGCTGCGAGGCTTCCCTGCTGTATCT 490	
Db	139227	CCACCTGCTGCGCCCGCCAGCTGCTGCAATTTCTAGTTGCTGCGAGGCTTCCCTGCTGTATCT 139286	
QY	491	CTAGCTGTTGCAAAACCCAGCTGCTGCCAGCACCACTGCTGCGCCCGCCAGCTGCTGTATCT 550	
Db	139287	CTAGCTGTTGCAAAACCCAGCTGCTGCCAGCACCACTGCTGCGCCCGCCAGCTGCTGTATCT 139346	
QY	551	CCAGCTGCTACAGGCGCCAGCTGCTGCCAGCCCTCTCTGCTGCGCCCGCGGCTTCTGCTCATTT 610	
Db	139347	CCAGCTGCTACAGGCGCCAGCTGCTGCCAGCCCTCTCTGCTGCGCCCGCGGCTTCTGCTCATTT 139406	
QY	611	CTAGTTGCTGTCATCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 670	
Db	139407	CTAGTTGCTGTCATCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 139466	
QY	671	CCACCTGCTGTAGAACCACTGCTCCACCCCATCTGCTGCGGCGAGTTCTTCTGCTGCTGAG 730	
Db	139467	CCACCTGCTGTAGAACCACTGCTCCACCCCATCTGCTGCGGCGAGTTCTTCTGCTGCTGAG 139526	
QY	731	TGAAGCTGCTCTGGATTTGCTGCAACCTTCTGCTCAACCTTTCAGTTTCAGGACACAGTA 790	
Db	139527	TGAGCTGCTCTGGATTTGCTGCAACCTTCTGCTCAACCTTTCAGTTTCAGGACACAGTA 139586	
QY	791	TCTATTTCAGAACATGTGGACTTCTGATGCTGTAAGAACAGGAGCATGGACTGATTTG 850	
Db	139587	TCTATTTCAGAACATGTGGACTTCTGATGCTGTAAGAACAGGAGCATGGACTGATTTG 139646	
QY	851	GAATAATTTTATTAGTATGATCTCTCTTTTATAGAGTTTTTATTCCTTATTGAATCTGA 910	
Db	139647	GAATAATTTTATTAGTATGATCTCTCTTTTATAGAGTTTTTATTCCTTATTGAATCTGA 139706	
QY	911	ATTACAGTCAAAATCCACATCATGTTTATAGAAATCTTATTTCTTAATTCATATACAT 970	
Db	139707	ATTACAGTCAAAATCCACATCATGTTTATAGAAATCTTATTTCTTAATTCATATACAT 139766	
QY	971	AAATCTCAATAGTATCTCTCTAGATGTTTCTTCTTAATGTTTCTGCTGATCAATTT 1030	
Db	139767	AAATCTCAATAGTATCTCTCTAGATGTTTCTTCTTAATGTTTCTGCTGATCAATTT 139826	
QY	1031	TCATGTGGAATGTTTGTATGTTCTCAATAAATCTTCATAGTGTTCAAAGCAAAAAA 1090	
Db	139827	TCATGTGGAATGTTTGTATGTTCTCAATAAATCTTCATAGTGTTCAAAGCAAAAAA 139886	

RESULT 3

AC025904/c
LOCUS
DEFINITION Homo sapiens clone RP11-585G19, WORKING DRAFT SEQUENCE, 17
unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC025904
AC025904.2 GI:7382590
HTG: HTGS_PHASE1: HTGS_DRAFT.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174032)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-585G19
Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 174032)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, J.,
Boguslavskiy, L., Bouckhgalter, B., Brown, A., Burkett, G.,
Campomare, A., Cooke, P., Choepel, Y., Colangelo, M., Collins, S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glend, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaKocque, K., Lamazares, R., Landers, T., Lehorzky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,
Melchior, J., Mihova, T., Miranda, C., Mlenda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodor, J., Tirrell, A., Travers, M., Triggiani, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:7249428.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBC
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8441
Center clone name: 585_G19
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158771 bases at least Q40
Consensus quality: 167389 bases at least Q30
Consensus quality: 170660 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 172432; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1
* 2588: contig of 2588 bp in length
* 2689: gap of 100 bp
* 2689: contig of 3777 bp in length
* 6466: gap of 100 bp
* 6566: contig of 2281 bp in length

*	8847	8945:	gap of	100 bp
*	8947	13104:	contig of 4158 bp	in length
*	13105	13204:	gap of	100 bp
*	13205	19430:	contig of 6226 bp	in length
*	19431	19530:	gap of	100 bp
*	19531	25564:	contig of 6034 bp	in length
*	25665	25664:	gap of	100 bp
*	25665	31377:	contig of 5713 bp	in length
*	31378	31477:	gap of	100 bp
*	31478	39523:	contig of 8046 bp	in length
*	39524	39623:	gap of	100 bp
*	39624	47042:	contig of 7419 bp	in length
*	47043	47142:	gap of	100 bp
*	47143	55411:	contig of 8269 bp	in length
*	55412	55511:	gap of	100 bp
*	55512	66000:	contig of 10489 bp	in length
*	66001	66100:	gap of	100 bp
*	66101	74184:	contig of 8084 bp	in length
*	74185	74284:	gap of	100 bp
*	74285	86712:	contig of 12428 bp	in length
*	86713	86812:	gap of	100 bp
*	86813	103527:	contig of 16715 bp	in length
*	103528	103627:	gap of	100 bp
*	103628	120465:	contig of 16838 bp	in length
*	120466	120565:	gap of	100 bp
*	120566	145510:	contig of 24945 bp	in length
*	145511	145610:	gap of	100 bp
*	145611	174032:	contig of 28422 bp	in length

ORIGIN		Query Match Best Local Similarity Matches	66.8%; 95.8%; 772;	Score 735.6; Pred. No. 7..8e-118; Conservative 0;	DB 2: Mismatches 19;	Length 174032; Indels 15;	Gaps 1:
Qy	300	CTGCTGTATCTTCAGTCTGCTGCACAAACCACCTGCTGCAGGACCACCTGCTGCAAGCCAGCAG					359
Dd	25563	CTGCTGCAACCCAGCTGCTGCCAGACCACTTTGTTGCCAGGACCACCTGCTGCGCTGGTAG					25504
Qy	360	CTGCTGCATTTTCAGTTGCTGCAGGCCCTTCCTGCTGTATCTCCAGCTGCTGCAAACTCAG					419
Dd	25503	TTGCTGCATTTTCAGTTTCTGCAGGCCCTTCCTGCTGTATCTCCAGCTGCTGCAAACTCAG					25444
Qy	420	-----CTGCTGCAGGACCACCTGCTGCCGCCCCCAGCTGTGCTATTCTTTAG					464
Dd	25443	CTGCTGCCAGACCACCTGCTGCAGGACCACCTGCTGCCGCCCCCAGCTGTGCTATTCTTTAG					25384
Qy	465	TTGCTGCAGGCCCTTCCTGCTGTATCTCTTAGCTGTTTGCAAACCCAGCTGCTGCAAGCTCAG					524
Dd	25383	TTGCTGCAGGCCCTTCCTGCTGTATCTCTTAGCTGCTGCAAACCCAGCTGCTGCCAGAGCTCAG					25324
Qy	525	CTGCTGCCGCCCTTCCTGCTGTATCTCCAGCTGCTACAGGCCCCCAGCTGCTGCAAGCTCTC					584
Dd	25323	CTGCTGCCGCCCCCAGCTGCTGTATCTCCAGCTGCTACAGGCCCCCAGCTGCTGCAAGCTCTC					25264
Qy	585	CTGCTGCCGCCCTTCCTGCTGTATCTAGTTTCTCATCCAGCTGCTGTGTTCTGCTGCTGAG					644
Dd	25263	CTGCTGCCGCTCGCTTCCTGCTGTATCTAGTTTCTCATCCAGCTGCTGTGTTCTGCTGAG					25204
Qy	645	CTGCTGCTGCCCTTCCTGCTGTGCCCCGAGCACCTGCTGTAGAACAACCTGCTGTAAGCTAT					704
Dd	25203	CTGCTGCTGCCCTTCCTGCTGTGCCCCGAGCACCTGCTGTAGAACAACCTGCTGTAAGCTAT					25144
Qy	705	CTGCTGCCGCCCTTCCTGCTGTGAGTGAAGCTGCTCTGGATTGTGTCACCTTCTGCTCTC					764
Dd	25143	CTGCTGCCGCCCTTCCTGCTGTGAGTGAAGCTGCTCTGGATTGTGTCACCTTCTGCTCTC					25084
Qy	765	TCAACCTTTCAGTTTCAGGCACAGAGTATCTATTTCAGAGAACATGTGGACTTCTGATGTCG					824
Dd	25083	TCAACCTTTCAGTTTCAGGCACCTGAGTATCTATTTCAGAGAACATGTGGACTTCTGATGTCG					25024
Qy	825	TGAACACAGAGGATGAGCTGATTTCGGAAATAATTTTTATTAGTATGTATCTGTTTTATAA					884
Dd	25023	TGAACACAGAGGATGAGCTGATTTCGGAAATAATTTTTATTAGTATGTATCTGTTTTATAA					24964
Qy	885	GAGCTTTTATTCCTATTGTAATCTGAATTTACAGTCAAAATCCACATCACTGTTTACA					944
Dd	24963	GAGCTTTTATTCCTATTGTAATCTGAATTTACAGTCAAAATCCACATCACTGTTTACA					24904
Qy	945	ATTCCTTTATTCCTAAATATACATAAATCTTCAAATGGTATPCCTTCTAGATGTTTCTT					1004
Dd	24903	ACTCTTTATTCCTAAATATACATAAATCTTCAAATGGTATPCCTTCTAGATGTTTCTT					24844
Qy	1005	CCTAATGTTTTCCTGCTGATCAATTTTCATGTGGAAATGTTTGTAGTGTCTCTCAATAAAC					1064
Dd	24843	CCTAATGTTTTCCTGCTGATCAATTTTCATGTGGAAATGTTTGTAGTGTCTCTCTCAATAAAC					24784
Qy	1065	TTCTACATGTTTCAAAAACAAAAAA 1090					
Dd	24783	TTCTACATGTTTCAAAAACAAAAAA 24758					

BASE COUNT	vector_side:right	50722 a	34581 c	34553 g	52576 t	1600 others
------------	-------------------	---------	---------	---------	---------	-------------

[illegible]

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 133392)
Worley,K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 133392)
Worley,K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17973061.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHAG
Center clone name: CH230-115J19
----- Summary Statistics
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 70190 bases at least Q40
Consensus quality: 73775 bases at least Q30
Consensus quality: 75923 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 1166 1165: contig of 1165 bp in length
* 1265 1265: gap of unknown length
* 2463 2463: contig of 1198 bp in length
* 2563 2563: gap of unknown length
* 3746 3746: contig of 1183 bp in length
* 3846 3846: gap of unknown length
* 5575 5575: contig of 1729 bp in length
* 5676 5676: gap of unknown length
* 6717 6717: contig of 1042 bp in length
* 6817 6817: gap of unknown length
* 7936 7936: contig of 1119 bp in length
* 8036 8036: gap of unknown length
* 9603 9603: contig of 1567 bp in length
* 9604 9604: gap of unknown length
* 10789 10789: contig of 1086 bp in length
* 10790 10790: gap of unknown length
* 10899 10899: contig of 1008 bp in length
* 11997 11997: gap of unknown length
* 11998 11998: contig of 1387 bp in length
* 13485 13485: gap of unknown length
* 14726 14726: contig of 1242 bp in length
* 14727 14727: gap of unknown length
* 14826 14826: gap of unknown length
* 16427 16427: contig of 1323 bp in length
* 16150 16150: gap of unknown length
* 16250 16250: contig of 1304 bp in length
* 17553 17553: gap of unknown length
* 17654 17654: contig of 1410 bp in length
* 19064 19064: gap of unknown length
* 20630 20630: contig of 1467 bp in length
* 20730 20730: gap of unknown length
* 21842 21842: contig of 1112 bp in length
* 21942 21942: gap of unknown length
* 21843 21843: gap of unknown length
* 21943 21943: gap of unknown length
* 22998 22998: contig of 1056 bp in length
* 23098 23098: gap of unknown length
* 24162 24162: contig of 1064 bp in length
* 24262 24262: gap of unknown length
* 25305 25305: contig of 1043 bp in length
* 25405 25405: gap of unknown length
* 25406 25406: contig of 1529 bp in length
* 26935 26935: gap of unknown length
* 27035 27035: contig of 1114 bp in length
* 28148 28148: gap of unknown length
* 28248 28248: gap of unknown length
* 30210 30210: contig of 1962 bp in length
* 30310 30310: gap of unknown length
* 32555 32555: contig of 1945 bp in length
* 32556 32556: gap of unknown length
* 33390 33390: contig of 1035 bp in length
* 33490 33490: gap of unknown length
* 35023 35023: contig of 1533 bp in length
* 35123 35123: gap of unknown length
* 37023 37023: contig of 1900 bp in length
* 37123 37123: gap of unknown length
* 38394 38394: contig of 1271 bp in length
* 38494 38494: gap of unknown length
* 39994 39994: contig of 1500 bp in length
* 40094 40094: gap of unknown length
* 41494 41494: contig of 1400 bp in length
* 41594 41594: gap of unknown length
* 42896 42896: contig of 1302 bp in length
* 42996 42996: gap of unknown length
* 45169 45169: contig of 2173 bp in length
* 45170 45170: gap of unknown length
* 46744 46744: contig of 1475 bp in length
* 46844 46844: gap of unknown length
* 48508 48508: contig of 1664 bp in length
* 48609 48609: gap of unknown length
* 50785 50785: contig of 2177 bp in length
* 50885 50885: gap of unknown length
* 52950 52950: contig of 2065 bp in length
* 53050 53050: gap of unknown length
* 55142 55142: contig of 2092 bp in length
* 55242 55242: gap of unknown length
* 56339 56339: contig of 1297 bp in length
* 56640 56640: gap of unknown length
* 58282 58282: contig of 1643 bp in length
* 58382 58382: gap of unknown length
* 59577 59577: contig of 1195 bp in length
* 59677 59677: gap of unknown length
* 61727 61727: contig of 2050 bp in length
* 61827 61827: gap of unknown length
* 63172 63172: contig of 1345 bp in length
* 63272 63272: gap of unknown length
* 65192 65192: contig of 1920 bp in length
* 65292 65292: gap of unknown length
* 67342 67342: contig of 2050 bp in length
* 67442 67442: gap of unknown length
* 69603 69603: contig of 2161 bp in length
* 69703 69703: gap of unknown length
* 71798 71798: contig of 2095 bp in length
* 71898 71898: gap of unknown length
* 74031 74031: contig of 2133 bp in length
* 74131 74131: gap of unknown length
* 75543 75543: contig of 1412 bp in length
* 75643 75643: gap of unknown length
* 77487 77487: contig of 1844 bp in length
* 77587 77587: gap of unknown length
* 79773 79773: contig of 2186 bp in length
* 79873 79873: gap of unknown length
* 82213 82213: contig of 2340 bp in length
* 82313 82313: gap of unknown length
* 85054 85054: contig of 2741 bp in length
* 85154 85154: gap of unknown length
* 88032 88032: contig of 2878 bp in length
* 88132 88132: gap of unknown length
* 89578 89578: contig of 1446 bp in length

```


(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1611: contig of 1611 bp in length
* 1612 1711: gap of unknown length
* 1712 2877: contig of 1166 bp in length
* 2878 2977: gap of unknown length
* 2978 4061: contig of 1084 bp in length
* 4062 4161: gap of unknown length
* 4162 5787: contig of 1626 bp in length
* 5788 5887: gap of unknown length
* 5888 7018: contig of 1131 bp in length
* 7019 7118: gap of unknown length
* 7119 8380: contig of 1262 bp in length
* 8381 8480: gap of unknown length
* 8481 10104: contig of 1624 bp in length
* 10105 10204: gap of unknown length
* 10205 11292: contig of 1088 bp in length
* 11293 11392: gap of unknown length
* 11393 12737: contig of 1345 bp in length
* 12738 12837: gap of unknown length
* 12838 14444: contig of 1607 bp in length
* 14445 14544: gap of unknown length
* 14545 15575: contig of 1031 bp in length
* 15576 15675: gap of unknown length
* 15676 17781: contig of 2106 bp in length
* 17782 17881: gap of unknown length
* 17882 19517: contig of 1636 bp in length
* 19518 19617: gap of unknown length
* 19618 21140: contig of 1523 bp in length
* 21141 21240: gap of unknown length
* 21241 22878: contig of 1638 bp in length
* 22879 22978: gap of unknown length
* 22979 24285: contig of 1307 bp in length
* 24286 24385: gap of unknown length
* 24386 26508: contig of 2123 bp in length
* 26509 26609: contig of 1440 bp in length
* 26609 28048: contig of 1440 bp in length
* 28049 28148: gap of unknown length
* 28149 30094: contig of 1946 bp in length
* 30095 30194: gap of unknown length
* 30195 31495: contig of 1300 bp in length
* 31495 31594: gap of unknown length
* 31595 34604: contig of 3010 bp in length
* 34605 34704: gap of unknown length
* 34705 36524: contig of 1920 bp in length
* 36525 36724: gap of unknown length
* 36725 38320: contig of 1596 bp in length
* 38321 38420: gap of unknown length
* 38421 39858: contig of 1438 bp in length
* 39859 39859: gap of unknown length
* 39859 43450: contig of 3452 bp in length
* 43451 43550: gap of unknown length
* 43551 46112: contig of 2562 bp in length
* 46113 46212: gap of unknown length
* 46213 48632: contig of 2420 bp in length
* 48633 48732: gap of unknown length
* 48733 50717: contig of 1985 bp in length
* 50718 50817: gap of unknown length
* 50818 53619: contig of 2802 bp in length
* 53620 53719: gap of unknown length
* 53720 56601: contig of 2882 bp in length
* 56602 56701: gap of unknown length
* 56702 56751: contig of 3050 bp in length
* 56752 59851: gap of unknown length
* 59852 62752: contig of 2901 bp in length
* 62753 62852: gap of unknown length

* 62853 66264: contig of 3412 bp in length
* 66265 66364: gap of unknown length
* 66365 70565: contig of 4201 bp in length
* 70566 70665: gap of unknown length
* 70666 74401: contig of 3736 bp in length
* 74402 74501: gap of unknown length
* 74502 79005: contig of 4504 bp in length
* 79006 79105: gap of unknown length
* 79106 84082: contig of 4977 bp in length
* 84083 84182: gap of unknown length
* 84183 87867: contig of 3685 bp in length
* 87868 87967: gap of unknown length
* 87968 93319: contig of 5352 bp in length
* 93320 93419: gap of unknown length
* 93420 97202: contig of 3783 bp in length
* 97203 97302: gap of unknown length
* 97303 101781: contig of 4479 bp in length
* 101782 101881: gap of unknown length
* 101882 107479: contig of 5598 bp in length
* 107480 107579: gap of unknown length
* 107580 111355: contig of 3776 bp in length
* 111356 111455: gap of unknown length
* 111456 115389: contig of 3934 bp in length
* 115390 115489: gap of unknown length
* 115490 119864: contig of 4375 bp in length
* 119865 124964: contig of 5000 bp in length
* 124965 125064: gap of unknown length
* 125065 129776: contig of 4712 bp in length
* 129777 129876: gap of unknown length
* 129877 138427: contig of 8551 bp in length
* 138428 138527: gap of unknown length
* 138528 145489: contig of 6962 bp in length
* 145490 145589: gap of unknown length
* 145590 154737: contig of 9148 bp in length
* 154738 154837: gap of unknown length
* 154838 164787: contig of 9950 bp in length
* 164788 164887: gap of unknown length
* 164888 173904: contig of 9017 bp in length
* 173905 174005: gap of unknown length
* 174006 182806: contig of 8802 bp in length.

FEATURES

Location/Qualifiers
1..182806

Query Match 36.6%; Score 403.4; DB 2; Length 182806;
Best Local Similarity 74.8%; Pred. No. 3.7e-60;
Matches 539; Conservative 0; Mismatches 166; Indels 16; Gaps 2;
QY 6 AACCCAGCTGAGAACTCCACGCTCTGACGCCATGCTCAGCTCCTGCTGTGGTCTGCTG 65
DB 57714 AACTCATCCAGAACCTCCA-CCTCTGACACCATGCTCAGCTCCTGCTGTGGTCTGCTG 57656
QY 66 CTCTGACCAGAGCTGTGTGTCAGGTCTCGGCCAGAGAGCTGCTGCCGCCAGCTGTG 125
DB 57655 CTCTGAGGAGGCTGTGCGCAAG-----GCTGCTGCCAGCCAGCTGCTG 57611
QY 126 CCAGACCACTTCTGCTCAGGACCACTGCTGCGGCCAGCTGCTGCTGCTGCTGCTG 185
DB 57610 CCAGACCACTTCTGCTCAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57551
QY 186 CAGGCTTCTGCTGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
DB 57550 CAGACCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57491
QY 246 CAGGACCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 305
DB 57490 CAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57431
QY 306 TATCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 365
DB 57430 CCGGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57371
QY 366 CATTTCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 425

```
Db 57370 CATTCTAGCTGCTGCAGGCTTCTCTGTCGCCGCCACCTGTTGCATTTCTAGCTGCTG 57311
QY 426 CAGGACCACTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 485
Db 57310 TAGGCTTCTGCTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57251
QY 486 TATCTCTAGCTGTTGCTGCAAAACCCAGCTGCTGCGACACCACTGCTGCGCCGCCAGCTGCTG 545
Db 57250 CCGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57191
QY 546 TATCTCCAGCTGCTACAGGCCCCAGCTGCTGCGACGCCCTGCTGCTGCTGCTGCTGCTGCTG 605
Db 57190 CATTCTAGCTGCTGCAGGCTTCTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTG 57131
QY 606 CATTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 665
Db 57130 CAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57071
QY 666 CCGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 725
Db 57070 CCGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57011
QY 726 C 726
Db 57010 C 57010

RESULT 7
AC067715
LOCUS
DEFINITION Mus musculus chromosome 11 clone RP23-326H14, WORKING DRAFT
ACCESSION AC067715
VERSION AC067715.24 GI:12025592
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Metzker, M.L., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Dederich, D., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Addison, S., Pace, A., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Buhay, C., Bunac, C., Davis, C., Delgado, D., Ding, Y., Dugan-Rocha, S.,
Cox, C., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Oguh, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 192027)
Worley, K.C.
Direct Submission
Submitted (27-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:11192108.
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: MACE
Center clone name: RP23-326H14
```

```
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 57% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 189676 bases at least Q40
Consensus quality: 191575 bases at least Q30
Consensus quality: 192401 bases at least Q20
Estimated insert size: 189593; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 7.8x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 71311: contig of 71311 bp in length
* 71312 71411: gap of unknown length
* 71412 116268: contig of 44857 bp in length
* 116269 116368: gap of unknown length
* 116369 145312: contig of 28944 bp in length
* 145313 145412: gap of unknown length
* 145413 173610: contig of 28198 bp in length
* 173611 173710: gap of unknown length
* 173711 190868: contig of 17158 bp in length
* 190869 190968: gap of unknown length
* 190969 192027: contig of 1059 bp in length.
FEATURES
Location/Qualifiers
1..192027
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-326H14"
BASE COUNT 55026 a 41185 c 39554 g 55749 t 513 others
ORIGIN
Query Match 34.4%; Score 378.2; DB 2: Length 192027;
Best Local Similarity 76.7%; Pred. No. 8.7e-56;
Matches 475; Conservative 0; Mismatches 143; Indels 1; Gaps 1;
QY 2 TGGGAACCTACACCAAGCTTCCACCTCTGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61
Db 33817 TGTAAACCCATCCAGAACCTCCA-CCTCTGACACCATGGCCAACTCTCTGCTGCTGCTG 33875
QY 62 TCTGCTCTGACACAGCTGCTGCTCAAGCTCTCGGCCAGGAGAGCTGCTGCCGACCACTGCT 121
Db 33876 TCTGCTCTGAGGAGAGCTGCTGCCAAGGCTGCTGCCAGCCAGCTGCTGCCAGACCACT 33935
QY 122 GCTGCACAGACCACTGCTGCCAGCACTGCTGCCGCCAGCTGCTGCCGCCAGCTGCTGCTG 181
Db 33936 GCTGTAGGACCACTGCTGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33995
QY 182 GCTGAGGCTTCTGCTGCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCCCTGACACT 241
Db 33996 GCTGCCAGTCTGCTGCTGCCAGCCACTGCTGCTGCCGCCAGCTGCTGCTGCTGCTGCTG 34055
QY 242 GCTGCAGGACCACTGCTGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
Db 34056 GCTGCAGGACCACTGCTGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 34115
QY 302 GCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCCAGGACCACTGCTGCCGCCAGCT 361
Db 34116 GCTGTGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34175
QY 362 GCTGCAATTTGCAAGTCTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
Db 34176 GCTGTATTTGCAAGCTGCTGCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34235
```


AUTHORS TITLE JOURNAL

Ellington, A.
Direct Submission
Submitted (26-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 28, 2002 this sequence version replaced gi:18135257.

COMMENT

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BM142E7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 192882 bases at least Q40
Consensus quality: 192997 bases at least Q30
Consensus quality: 193039 bases at least Q20
Insert size: 197658; sum-of-contigs
Insert size: 203092; 2.3% error; agarose-fp
Quality coverage: 13.20x in Q20 bases; sum-of-contigs Quality
coverage: 13.17x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES source

1..197658
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-142E7"
/clone_lib="RPC1-23"
misc_feature
1..197658
/note="assembly_fragment:02832"
BASE COUNT 59207 a 42073 c 41216 g 55162 t
ORIGIN

Query Match 33.4%; Score 368.2; DB 2; Length 197658;
Best Local Similarity 72.7%; Pred. No. 4.8e-54;
Matches 524; Conservative 0; Mismatches 178; Indels 19; Gaps 3;

Oy 6 AACCCACCAGAACCTCCACCCCTGCACGCCAATGCGTACGCTCTCTGCTCTCTG 65
||||| ||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17709 AACCTACCAAGAACCTTACCC-TCTAAACCATGGTCAGCTCCTGTGTGCTCTCTG 17767
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 66 CTCTGACCAGAGCTGTGTCAAGCTCTGCGCCAGGAGAGCTGTCGCCGCCAGCTGCTG 125
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17768 CTCTGAGGAGGGCTGTGGCCAAAG-----GCTGCTGCCAGCCAGCTGCTG 17812
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 126 CCAGACCACTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCATTTCCAGTTGCTG 185
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17813 CCAGACTACTGCTGTAGGACCACTGCTGCCGCCAGCTGCTGTGTGCTGCTGCTG 17872
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 186 CAGGCCCTCTCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCACTGCTG 245
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17873 CAGGCCCACTGCTGTGTCTCCAGCTGCTGCAGACCCAGCTGTTGCCAGTCTGTGCTG 17932
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 246 CAGGACCACTGCTGCCGCCAGCTGCTGCATTTCCAGTTGCTGAGGCCCTTCTGCTG 305
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17933 CCAGCCCACTGCTGCCGCCAGCTGCTGCCGCCAGCTGCTGCTGCTGCTGCTG 17992
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 306 TATCTCCAGTGTGCAAAACCCAGCTGCTGCAGACCACTGCTGCCGCCAGCTGCTG 365
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17993 TAGACCTTGTGCTGCGC---CCCAGCTGCTGTGTGCTGCTGCTGCTGCTGCTG 18049
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 366 CATTTCCAGTGTGCTGAGGCTCTCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTG 425
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18050 CCAGTCTGTGCTGCCAGCCCACTGCTGCCGCCAGCTGCTGTGCTGCCCACTGTTG 18109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 426 CAGGACCACCTGCTGCCGCCAGCTGCTGCATTTCTAGTTGCTGCAGGCCCTTCTGCTG 485
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18110 CATTTCCAGCTGCTGCCGCCCTGCTGTGTAGTTCCAGCTGCTGTGGATCTAGCTGCTG 18169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 486 TATCTCTAGCTGTTGCAAAACCCAGCTGCTGCCAGACCACTGCTGCCGCCAGCTGCTG 545
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18170 CGGCCCACTGCTGCAATTTCCAGCTGCTGCCGCCAGCTGCTGCCAGCCAGCTGCTG 18229
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 546 TATCTCCAGCTGCTACAGGCCCACTGCTGCCAGCCCTCTGCTGCCGCCAGCTGCTGCTG 605
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18230 TGTGTCCAGCTGCTGCAGACCCCACTGCTGCATCTCCAGCTGCTGCCGCCAGCTGCTG 18289
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 606 CATTTCTAGCTGCTGCATGCCAGCTGCTGTGTCCAGCTGCTGCCGCCCTTTCCAGCTG 665
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18290 TAGTTCTAGCTGCTGTGGATCTAGCTGCTGCCGCCCACTGCTGCCAGCCCACTGCTG 18349
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 666 CCCGACCACCTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCTGCCAGCTGCTGCTG 725
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18350 TGTGTCCAGCTGCTGCAGACCCCACTGCTGCTCATCTCTAGCTGCTGCCGCCAGCTGCTG 18409
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 726 C 726
Db 18410 C 18410

RESULT 10 AL590997 LOCUS

DEFINITION Mouse DNA sequence from clone RP23-132J20 on chromosome 11,
complete sequence.

ACCESSION AL590997
VERSION AL590997.13 GI:17221205

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 200602)

Blakey, S.

Direct Submission

Submitted (29-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Nov 30, 2001 this sequence version replaced gi:16973924.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-132J20 is
from the RP23-23 Mouse PAC Library
constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

This sequence is the entire insert of clone RP23-132J20 The true

left end of clone RP23-212C14 is at 95441 in this sequence. The

true right end of clone RP23-272C1 is at 89011 in this sequence.

Location/Qualifiers

1..200602

/organism="Mus musculus"

FEATURES source

```
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-132J20"
/clone_11b="RPC1-23"
42963
/note="Random repeat. Forced join. Gap size estimated to
be approximately 260bp by restriction digest data."
misc_feature 115069..115073
/note="1327 bases of IS2 transposon (V0061) removed here.
Flanking repeat sequences are irregular."
misc_feature 115074..115111
/note="Sequence from overlapping clone RP23-212C14
(AL590992). Assembly confirmed by restriction digest."
misc_feature 189825..189863
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
BASE COUNT 58430 a 40464 c 42904 g 58804 t
ORIGIN

Query Match 33.4%; Score 368.2; DB 10; Length 200602;
Best Local Similarity 72.7%; Pred. No. 4.8e-54;
Matches 524; Conservative 0; Mismatches 178; Indels 19; Gaps 3;

Oy 6 AACCCACCCAGAACCTCCACCTCTGACGCCATGCTCAGCTCCTGCTGCTGCTGCTG 65
Db 174439 AACCTACAGAACCTTCACC-TCATAAACATGCTCAGCTCCTGTTGTGCTGCTG 174497

Oy 66 CTCTGACAGAGCTGTGCTGAAGTCTCGGCCAGGAGAGCTGCTGCGCCGCCAGCTGCTG 125
Db 174498 CTCTGAGAGAGGCTGTGGCCAAAG-----GCTGCTGCCAGCCAGCTGCTG 174542

Oy 126 CCAGACCACTGCTGCGAGGACACCTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTG 185
Db 174543 CCAGACTACCTGCTGTAGGACCACTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTG 174602

Oy 186 CAGGCTTCTCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCACTGCTG 245
Db 174603 CAGGCCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174662

Oy 246 CAGGACCACTGCTGCGCCGCCAGCTGCTGATTTTCAGTGTGCTGCGAGCCCTTCTGCTG 305
Db 174663 CCAGCCCACTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174722

Oy 306 TATCTCCAGCTGCTGCAAAACCCAGCTGCTGCGAGACCACTGCTGCGCCGCCAGCTGCTG 365
Db 174723 TAGACTTGTGCTGCGC---CCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174779

Oy 366 CATTTCCAGTGTGCTGCGAGGCTTCTGCTGTATCTCCAGCTGCTGCTGCAAAACCCAGCTGCTG 425
Db 174780 CCAGTCTGTGCTGCGCAGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174839

Oy 426 CAGGACCACTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 485
Db 174840 CATTTCCAGCTGCTGCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174899

Oy 486 TATCTCTAGCTGTGCTGCAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 545
Db 174900 CCGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174959

Oy 546 TATCTCCAGCTGCTACAGGCCCCAGTGTGCTGCGAGCCCTTCTGCTGCTGCTGCTGCTGCTG 605
Db 174960 TGTGTCCAGCTGCTGCGACACCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175019

Oy 606 CATTTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 665
Db 175020 TAGTTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175079

Oy 666 CCGGACCACTGCTGTAGAAACCTGCTTCCACCCCACTGCTGCTGCTGCTGCTGCTGCTG 725
Db 175080 TGTGTCCAGCTGCTGCGACACCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175139

Oy 726 C 726
|
```

Db 175140 C 175140

RESULT 11
AC025904
LOCUSDEFINITION Homo sapiens clone RP11-585G19, WORKING DRAFT SEQUENCE, 17
unordered pieces.

ACCESSION AC025904.2 GI:7382590

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 174032)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-585G19

Unpublished

2 (bases 1 to 174032)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavsky,I., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,C., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,

Klein,J., Lacroque,K., Lamazares,R., Landers,T., Leloczky,J.,

Levine,K., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPherson,K.,

Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienya,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Spencer,B.,

Roy,A., Santos,R., Schauer,S., Severy,P., Severi,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triplino,J.,

Vassiliev,V., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,P., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 1, 2000 this sequence version replaced gi:7249428.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8441

Center clone name: 585.G.19

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158771 bases at least Q40

Consensus quality: 167389 bases at least Q30

Consensus quality: 170660 bases at least Q20

Insert size: 170000; agarose-fp

Quality coverage: 3.9 in Q20 bases; agarose-fp

Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 17 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2588: contig of 2588 bp in length
* 2589 2688: gap of 100 bp
* 2689 6465: contig of 3777 bp in length
* 6466 6565: gap of 100 bp
* 6566 8845: contig of 2281 bp in length
* 8847 8946: gap of 100 bp
* 8947 13104: contig of 4158 bp in length
* 13105 13204: gap of 100 bp
* 13205 19430: contig of 6226 bp in length
* 19431 19530: gap of 100 bp
* 19531 25564: contig of 6034 bp in length
* 25565 25664: gap of 100 bp
* 25665 31377: contig of 5713 bp in length
* 31378 31477: gap of 100 bp
* 31478 39523: contig of 8046 bp in length
* 39524 39623: gap of 100 bp
* 39624 47042: contig of 7419 bp in length
* 47043 47142: gap of 100 bp
* 47143 55411: contig of 8269 bp in length
* 55412 55511: gap of 100 bp
* 55512 66000: contig of 10489 bp in length
* 66001 66100: gap of 100 bp
* 66101 74184: contig of 8084 bp in length
* 74185 74284: gap of 100 bp
* 74285 86712: contig of 12428 bp in length
* 86713 86812: gap of 100 bp
* 86813 103527: contig of 16715 bp in length
* 103528 103627: gap of 100 bp
* 103628 120465: contig of 16838 bp in length
* 120466 120565: gap of 100 bp
* 120566 145510: contig of 24945 bp in length
* 145511 145610: gap of 100 bp
* 145611 174032: contig of 28422 bp in length.

FEATURES

Source
1. 174032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-585G19"
/clone_lib="RP11-11 Human Male BAC"
misc_feature
1. 2588
/note="assembly_fragment"
vector_side:right
2689..6465
/note="assembly_fragment"
6566..8846
/note="assembly_fragment"
8947..13104
/note="assembly_fragment"
13205..19430
/note="assembly_fragment"
19531..25564
/note="assembly_fragment"
25665..31377
/note="assembly_fragment"
31478..39523
/note="assembly_fragment"
39624..47042
/note="assembly_fragment"
47143..55411
/note="assembly_fragment"
55512..66000
/note="assembly_fragment"
66101..74184
/note="assembly_fragment"
74285..86712
/note="assembly_fragment"
86813..103527
/note="assembly_fragment"
103628..120465

misc_feature
/note="assembly_fragment"
120566..145510
/note="assembly_fragment"
145611..174032
/note="assembly_fragment"
clone_end:sp6
vector_side:right
BASE COUNT 50722 a 34581 c 34553 g 52576 t 1600 others
ORIGIN
Query Match 32.3%; Score 356; DB 2; Length 174032;
Best Local Similarity 69.4%; Pred. No. 6.2e-52;
Matches 504; Conservative 0; Mismatches 210; Indels 12; Gaps 1;
Qy 1 CTGGGAACCCAGCCAGAACCTCTACACCTCTGACCCCATGTGTCAGCTCTCTGCTGATCTCT 60
Db 6672 CTGGAAACCCAGCCAGAACCTCTACACCTCTGACACCATGTGTCAGCTCTCTGCTGATCTCT 6731
Qy 61 GTCGCTCTGACACAGAGCTGTGGTCAAGGTCGCGCCAGGAGAGCTGTGCGGCGGAGC 120
Db 6732 GTGTGCTCTGACACAGAGCTGT-----GGCCTAGAGAACTGCTGTGCTGCTGACG 6779
Qy 121 TGCTGCCAGAGCCAGCTGTGTCAGGACCACTGTGTCGCCGCCAGCTGCTGTGATTTGCAAT 180
Db 6780 TGCTGCCAGAGCCAGCTGTGTCAGGACCACTGTGTCGCCGCCAGCTGCTGTGCTGTCAGC 6839
Qy 181 TGCTGCCAGGCTTCTCTGCTGTATCTCCAGCTGTGTCAGAACCCAGCTGCTGTGCTGATGAC 240
Db 6840 TGCTGCCAGGCTTCTCTGCTGTATCTCCAGCTGTGTCAGAACCCAGCTGCTGTGCTGATGAC 6899
Qy 241 TGCTGCCAGGACCACTGTGTCGCCGCCAGCTGTGTCGATTTCCAGCTGCTGTGCAAGCTTCC 300
Db 6900 TGCTGTGTCGACCTGTGTCAGACCCCACTGTGTCGACCTGTGTCGCTGCTGCTGCTGCTGAC 6959
Qy 301 TGCTGTATCTCCAGCTGTGTCAAACCCAGCTGTGTCAGGACCACTGTGTCGCGGCGGAGC 360
Db 6960 TGCTGCCAGACCCCACTGTGTCGACCACTGTGTCGCTGTAGGACCACTGTGTCGCGGCGGAGC 7019
Qy 361 TGCTGCATTTCCAGTTGCTGTCAGGCTTCTCTGCTGTATCTCCAGCTGCTGTGCAAGCTTCC 420
Db 7020 TGCTGCCAGGCTTCTCTGCTGTGTCGCTGTGTCGCTGTCGACCACTGTGTCGCTGCTGCTG 7079
Qy 421 TGCTGCCAGGACCACTGTGTCGCCGCCAGCTGTGTCGATTTCCAGCTGCTGTGCTGCTGCTG 480
Db 7080 TACTGTGTCGACCTGTGTCAGACCCCACTGTGTCGACCACTGTGTCGCTGCTGCTGCTGCTG 7139
Qy 481 TGCTGTATCTCTAGCTGTGTCAAACCCAGCTGTGTCGACCACTGTGTCGCGGCGGAGC 540
Db 7140 TGCTGCCGCGGCGGAGCTGTGTCGCTGCTGTCGCTGCTGTCGCTGCTGCTGCTGCTGCTGCT 7199
Qy 541 TGCTGTATCTCCAGCTGCTAGAGGCTGCTGTCGACGCTGCTGTCGCGGCTGCTGCTGCTGCTG 600
Db 7200 TGCTGTAGGCTGCTGCTGTCGCAACCACTGTGTCGACGACCACTGTGTCGCTGCTGCTGCT 7259
Qy 601 TGCTGCATTTCTAGCTGCTGTCATCCAGCTGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 7260 TATTGCATTTCTAGCTGCTGTCGACCTTCTCTGACGATATCTCTAGCAGTAGTAGTCTGCTG 7319
Qy 661 AGCTGCCGCGGACCACTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCTGCGGCGGAGC 720
Db 7320 TGCCTGCTGCTGCTGCTGTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7379
Qy 721 TGCTGC 726
Db 7380 TGCTGC 7385
RESULT 12
AC107303/c AC107303 347572 bp DNA linear HTG 09-MAY-2002
LOCUS Homo sapiens chromosome 3 clone RP11-595122, WORKING DRAFT
DEFINITION SEQUENCE, 32 unordered pieces.
ACCESSION AC107303

AC107303.3 GI:20335965
 HTG: HTGS_PHASE1: HTGS_DRAFT.
 SOURCE
 Homo sapiens
 ORGANISM

REFERENCE
 AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buay,C., Burch,P., Burrett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escott,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Hombi,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H., Lozador,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nkokenko,S., Ogih,M., Okuwono,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

Direct Submission
 Unpublished
 Worley,K.C.
 2 (bases 1 to 347572)
 Direct Submission
 Submitted (18-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 347572)
 Direct Submission
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Apr 28, 2002 this sequence version replaced gi:18449841.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HDSS
 Center clone name: RP11-595I22
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329
 Consensus quality: 331666 bases at least Q40
 Consensus quality: 339305 bases at least Q30
 Consensus quality: 344140 bases at least Q20
 Estimated insert size: 172965; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 2351: contig of 2351 bp in length
 2352: gap of unknown length
 2452: contig of 3741 bp in length
 6193: gap of unknown length
 6293: contig of 4047 bp in length
 10340: gap of unknown length
 10440: contig of 3242 bp in length
 13682: gap of unknown length
 13782: contig of 4512 bp in length
 18294: gap of unknown length
 18394: contig of 4541 bp in length
 22935: gap of unknown length
 23035: contig of 3571 bp in length
 26605: gap of unknown length
 26706: contig of 4925 bp in length
 31631: gap of unknown length
 31731: contig of 5267 bp in length
 36998: gap of unknown length
 37098: contig of 4534 bp in length
 41631: gap of unknown length
 41732: contig of 6721 bp in length
 48453: gap of unknown length
 48553: contig of 7254 bp in length
 55806: gap of unknown length
 55907: contig of 9282 bp in length
 65189: gap of unknown length
 65289: contig of 6569 bp in length
 71858: gap of unknown length
 80512: contig of 8555 bp in length
 80612: gap of unknown length
 87649: contig of 7037 bp in length
 87750: gap of unknown length
 94122: contig of 6373 bp in length
 94222: gap of unknown length
 103064: contig of 8842 bp in length
 103065: gap of unknown length
 103165: contig of 8895 bp in length
 112059: gap of unknown length
 112160: contig of 8725 bp in length
 120884: gap of unknown length
 120885: contig of 13174 bp in length
 134158: gap of unknown length
 134258: contig of 12573 bp in length
 146831: gap of unknown length
 146931: contig of 11282 bp in length
 158213: gap of unknown length
 158214: contig of 13888 bp in length
 172201: gap of unknown length
 172202: contig of 11927 bp in length
 184228: gap of unknown length
 184329: contig of 11170 bp in length
 195498: gap of unknown length
 195598: contig of 15148 bp in length
 210746: gap of unknown length
 210847: contig of 15482 bp in length
 226428: gap of unknown length
 226329

226429	*	246181:	contig of 19753 bp in length	
246182	*	246281:	gap of unknown length	
246282	*	265510:	contig of 19229 bp in length	
265511	*	265610:	gap of unknown length	
265611	*	293686:	contig of 28076 bp in length	
293687	*	293786:	gap of unknown length	
293787	*	347572:	contig of 53786 bp in length.	
FEATURES				
Location/Qualifiers				
1..347572				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/chromosome="3"				
/clone="RP11-595122"				
BASE COUNT	107326	a	64859	c 63333 g 108900 t 3154 others
ORIGIN				
Query Match 32.3%; Score 356; DB 2: Length 347572;				
Best Local Similarity 69.4%; Pred. No. 6.5e-52;				
Matches 504; Conservative 0; Mismatches 210; Indels 12; Gaps 1;				
Qy	1	CTGGAACCCACCCAGAACCTCCACCCCTCTGACGCCATGTGTCAGCTCTGCTGGCTCT	60	
Db	12544	CTGGAACCCACCCAGAACCTCCACCCCTCTGACACCATGTGTCACCTCTGTGTGGCTCT	12485	
Qy	61	GTCTGCTGTGACAGAGCTGTGGTCAAGGTCTGGCGCAGGAGAGCTGTGCGGCCCCAGC	120	
Db	12484	GTGTGCTGTGACAGGGTGT-----GGCCTAGAGAACTGTGCGCTCCAGC	12437	
Qy	121	TGCTGCCAGACCACTGCTGTCAGGACCACTGCTGCCGCCACAGCTGCTGATTTCCAGT	180	
Db	12436	TGCTGCCAGACCACTGCTGTCAGGACCACTGCTGCCGCCACAGCTGCTGTTGTCAGC	12377	
Qy	181	TGCTGACGACCTTCTCTGTATATCTCCAGCTGCTGCAAAACCCAGTGTGCTGACCAAC	240	
Db	12376	TGCTGACGACCACTGCTGTCAGACCACTGCTGTCAGACCACTGCTGCCACCCAGC	12317	
Qy	241	TGCTGACGACCACTGCTGTCGCCGCCACAGCTGCTGCATTTCCAGTGTGTCAGGCTTCC	300	
Db	12316	TGCTGTGTGTCAGCTGCTGTCAGACCCAGCTGCTGCCAGTGTGTGTCGACGCCAC	12257	
Qy	301	TGCTGTATCTCAGCTGCTGCAAAACCCAGCTGCTGACGACCACTGCTGCCGCCCCAGC	360	
Db	12256	TGCTGACGACCACTGCTGTCAGGACCACTGCTGTCAGGACCACTGCTGCCGCCCCAGC	12197	
Qy	361	TGCTGATTTCCAGTTGCTGTCAGGCTTCCCTGTATCTCCAGCTGCTGCAAAACCCAGC	420	
Db	12196	TGCTGACGCCCACTGCTGTCAGGCTGCTGTCAGGCTGCTGTCAGGCTGCTGTCAGGCT	12137	
Qy	421	TGCTGACGACCACTGCTGTCGCCGCCACAGCTGCTGCATTTCCAGTGTGTCAGGCTTCC	480	
Db	12136	TACTGTGTGTCAGCTGCTGTCAGACCCAGCTGCTGCCAGACCACTGCTGTCAGAACCA	12077	
Qy	481	TGCTGTATCTTAGCTGTGCAAAACCCAGCTGCTGCCAGACCACTGCTGCCGCCCCAGC	540	
Db	12076	TGCTGCCGCCCACTGCTGTGTCAGCTGCTGTCAGGCTGCTGTCAGGCTGCTGTCAGCT	12017	
Qy	541	TGCTGTATCTCAGCTGTACAGGCCCACTGCTGCCAGCCCTGCTGTCGCCGCCCGCT	600	
Db	12016	TGCTGTAGCCATCTGCTGTCGCAAAACCACTGCTGTCAGACCACTGCTGCCACCCAGC	11957	
Qy	601	TGCTGATTTTCTAGTTGCTGTCATCCCACTGCTGTCAGCTGCTGCCGCCCTTTC	660	
Db	11956	TATTGATTTTCCAGCTGCTGCCAGCTTCTTCCAGTATCTTACGACAGTAGTCTCTCC	11897	
Qy	661	AGCTGCCCAACCACTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCCGCCAGTTCT	720	
Db	11896	TGCTGTGGCTTCTAGCTGCTGACAGGCTTCTTCTGCTCATCTCCAGTTGCTGCCGCCAAC	11837	
Qy	721	TGCTGC 726		
Db	11836	TGCTGC 11831		

```
Db 233 TGCTGTGTGTCAGCTGCTGCAGACCCAGTGTGTCAGTCTGTGTGCTGCCAGCCACC 292
Qy 301 TGCTGTATCTCCAGCTGCTGCAAAACCAGTGTGTGCAGACCCAGCTGCTGCCGCCCCAGC 360
Db 293 TGCTGCAGACCCCAATGCTGCCAGACTACCTGTGTGTAGGACCACTGCTGCCGCCCCAGC 352
Qy 361 TGCTGCATTTCCAGTGTGCTGCAGGCCCTTCTGCTGTATCTTCCAGTGTGCTGCAAAACCAGC 420
Db 353 TGCTGCAGGCCCCAGTGTGCTGCCAGTCTGTGTGCTGCCAGCCCACTGCTGCTGCCCCAGC 412
Qy 421 TGCTGCAGGACCACTGCTGCCGCCCCAGTGTGTCATTTCTAGTTGCTGCAGGCCCTTC 480
Db 413 TACTGTGTGTCAGCTGCTGCAGACCCCACTGCTGCCAGACCACTGCTGTCAGAAACCAAC 472
Qy 481 TGCTGTATCTAGCTGTGCAAAACCAGTGTGCTGCCAGACCACTGCTGCCGCCCCAGC 540
Db 473 TGCTGCCGCCCCAGCTGCTGTGTGTCAGAGTGTGTACAGGCCCACTTGTGCCCACTCTTA 532
Qy 541 TGCTGTATCTCCAGCTGCTACAGGCCCACTGCTGCCAGGCCCTCTGCTGCCGCCCCGGCT 600
Db 533 TGCTGTAGCCATCTGCTGGCAAAACCACCTGCTACAGGACCACTGCTGCCACCCAGC 592
Qy 601 TGCTGCATTTCTAGTGTGTCATCCAGTGTGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTC 660
Db 593 TATTGCATTTCCAGTGTGTCAGGCCCTTCTGTCAGTATCTCTAGCAGCAGTAGTCTCCTCC 652
Qy 661 AGCTGCCGCCACCACTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCTGCCAGTCTCT 720
Db 653 TGCTGTGGCTTCAGCTGCTGCAGGCTCTCTGCTGTCATCTCCAGTGTGCTGCCGCCCAAC 712
Qy 721 TGCTGC 726
Db 713 TGCTGC 718

RESULT 14
LOCUS HSA406939 1203 bp mRNA linear PRI 31-JAN-2001
DEFINITION Homo sapiens mRNA for keratin associated protein 4.7 (KRTAP4.7
gene).
ACCESSION AJ406939
VERSION AJ406939.1 GI:12655451
KEYWORDS keratin associated protein 4.7; KRTAP4.7 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1203)
Rogers, M.A., Langbein, L., Winter, H., Elmann, C., Korn, B. and
Schweizer, J.
Characterization of a cluster of human high/ ultrahigh keratin
associated proteins on chromosome 17q12-21
Unpublished
JOURNAL REFERENCE 2 (bases 1 to 1203)
Rogers, M.A.
Direct Submission
TITLES Submitted (16-OCT-2000) Rogers M.A., Research Program B, German
Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg,
Germany 69120, GERMANY
JOURNAL Related genomic sequence: AC025904 (142457-143089nt, neg strand).
FEATURES
Location/Qualifiers
1..1203
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q12-21"
/clone="sc1181"
/tissue_type="scalp"
47..679
/gene="KRTAP4.7"
47..679
/gene="KRTAP4.7"
/function="structural protein"
```

```
polyA_signal 242 a 389 c 248 g 324 t
BASE COUNT 1127.11132
ORIGIN
Query Match 32.2% Score 354.4 DB 9; Length 1203;
Best Local Similarity 71.0% Pred. No. 7.9e-52;
Matches 491; Conservative 0; Mismatches 186; Indels 15; Gaps 1;
Qy 2 TGGAGACCCACCCAGAACCTCCACCCCTCTGACGCCATGTCAGTCTCTGTCGATCTG 61
Db 12 TGAAGACCCACCCAGATCTCCCGTTCGACACCATGGTCAGTCTCTGTCGATCTG 71
Qy 62 TCTGCTCTGACAGAGCTGTGTCAGGTCCTGGCCAGGAGAGCTGCTGCCGCGGCACT 121
Db 72 TGTGCTCTGACAGAGCTGTGTCAGGTCCTGGCCAGGAGAGCTGCTGCCGCGGCACT 131
Qy 122 GCTGCCAGACCACTGCTGTGCAGGACCACTGCTGCCGCCCCAGCTGCTGTCATTTCA 181
Db 132 GCTGTACAGACCACTGCTGTGCAGGACCACTGCTGCCGCCCCAGCTGCTGTCGAT 191
Qy 182 GCTGCCAGGCTTCTGCTGTATCTTCCAGCTGTCGCAAAACCAGCTGCTGTCATTTCA 241
Db 192 GCTGCCAGGCTTCTGCTGTATCTTCCAGCTGTCGCAAAACCAGCTGCTGTCGATTTCA 251
Qy 242 GCTGCCAGGACCACTGCTGCCGCCCCAGCTGTCGATTTCCAGTGTGTCGATTTCA 301
Db 252 GCTGTGAGAGGACCTGCTGCCACCACTAGTGTGTCATCTGCCAGTGTGTCGTCAG 311
Qy 302 GCTGTATCTCCAGCTGTCGCAAAACCAGCTGTCGCAAGACCACTGCTGCCGCCCCAG 361
Db 312 GCTGTATCTCCAGCTGTCGCAAGACCACTGTCGTCAGTGTGTCGTCGTCAGGCT 371
Qy 362 GCTGCATTTCCAGTGTGTCGAGGCTTCTCTGCTATCTCCAGTGTGTCGTCGTCGTC 421
Db 372 GCTGCCACCCAG-----CTGCTGCATCTCCAGCTGCTGTCGTCGTCGTCGTC 416
Qy 422 GCTGCCAGGACCACTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 481
Db 417 GCTGTGTGTCAGGCTGTCGAGGCCAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTC 476
Qy 482 GCTGTATCTCTAGTGTGTCGCAAAACCAGCTGTCGCCAGACCACTGCTGCCGCCCCAG 541
Db 477 GCTGCCGTCGAGCTGCTGTCATCTTCCAGCTGCTGCCGCCCCCTCTTGTGTGAATTC 536
Qy 542 GCTGTATCTCCAGTGTGTCAGGCCCCAGTGTGCCAGCCCTCTGCTGTCGTCGTCGTC 601
Db 537 GCTGCCGCCATGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 596
Qy 602 GCTGCATTTCTAGTGTGTCATCTCCAGCTGCTGTGTCGTCGTCGTCGTCGTCGTCGTC 661
Db 597 CTTGCCACACCACTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 656
Qy 662 GCTGCCGCCACCACTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 693
Db 657 GCTGTGCTCTCTCTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 688

RESULT 15
AC037482/2 AC037482 186078 bp DNA linear PRI 23-FEB-2002
LOCUS Homo sapiens chromosome 17, clone RP11-619M15, complete sequence.
DEFINITION AC037482
ACCESSION AC037482
VERSION AC037482.14 GI:18873872
KEYWORDS HTG.
```


Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	701.5	47.3	186	2	A45910	ultra-high-sulfur
2	693.5	46.7	223	2	B38346	ultra-high-sulfur
3	684	46.1	220	2	A38346	ultra-high-sulfur
4	619	41.7	151	2	S60314	hair keratin cysteine-rich hair
5	569.5	38.4	126	2	I46489	high-sulfur protein
6	560	37.7	188	2	JC6547	high-sulfur keratin
7	463.5	31.2	175	2	S37649	high-sulfur keratin
8	451	30.4	177	2	S37650	high-sulfur keratin
9	434	29.2	182	1	KRSHHD	keratin high-sulfur
10	432	29.1	182	2	I47105	high-sulfur wool m
11	431.5	29.1	169	1	S18946	ultra high-sulfur
12	408.5	27.5	172	1	KRSHHA	keratin high-sulfur
13	402	27.1	172	2	I47106	high-sulfur wool m
14	384.5	25.9	162	2	I47107	high-sulfur wool m
15	375	25.3	654	2	T30136	hypothetical prote
16	360.5	24.3	152	2	I47108	high-sulfur wool m
17	358.5	24.2	152	1	KRSHHC	keratin high-sulfur
18	358.5	24.2	152	2	I47111	high-sulfur wool m
19	358	24.1	251	2	A55035	cysteine-rich prot
20	350.5	23.6	152	2	I47109	high-sulfur wool m
21	350.5	23.6	152	2	I47112	high-sulfur wool m
22	348	23.5	156	1	KRSHHB	keratin high-sulfur
23	344	23.2	191	2	I46412	keratin KAP5.4 - s
24	325	21.9	197	2	I46413	keratin KAP5.5 - s
25	305.5	20.6	182	2	A36686	ultra-high-sulfur
26	298.5	20.1	572	2	T29880	hypothetical prote
27	296.5	20.0	131	1	KRGF3M	keratin high-sulfur
28	294	19.8	728	2	T20561	hypothetical prote
29	293	19.7	425	2	T18592	hypothetical prote

```
C:Superfamily: ultra-high-sulfur keratin
```

Query Match	38.4%:	Score 569.5;	DB 2;	Length 126;
Best Local Similarity	57.7%:	Pred. NO. 1.5e-33;		
Matches 90; Conservative 15; Mismatches 20; Indels 31; Gaps 4;				
Qy	1	MVSCCGSVCSAGUGLGO - ESCCRPSCQTTCRTCCRPSCCISCCRPSCCISSC 59		
Db	1	MVSCCGSVCSERGGUUFCEESCRPS-----CCRPCCQPSCCRPTCCISSC 50		
Qy	60	CKPSCCLLTCTTKTTCRPSVCISCCRPSCCISSCCRKPSCCRTTCRRPSCCISCCRPS 119		
Db	51	CRPOCCQSVCRQPTCCRPSCVSISSCRP-----TCCRPTCCRPCTCCRP----- 93		
Qy	120	CISSCKRPSCKTTKTCPSPSCISCCRPSCCISSCCRPSCCISSCC 155		
Db	94	---TSQCITCTCTGCTGCRPCSCVSTCCRPCCSSSGGCC 126		


```

Db      58  -----TSIQSCQPISIQNSCCOPTCLQTSGGCTGGCIGGSIQGVQSSGAVSSRTKW 111
QY      105  CRPSCCISSCCRPSCCISSCCKPSCC-----RTTCRPSCCISSCCRPSCCISSCCKPSC 159
      111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      113  CPDCCRVEGTSLPCCVPIQTSPPCCQLYYAQASCCRPSPYCGSGCCRPAC-----C 163
QY      160  CRTTCRPSCCISSCYRPOCCOPSC 184
      111 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      164  COPTCIEPVC-----EPTCCEPTC 182

RESULT 10
147105
high-sulfur wool matrix protein B2A - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 15-Sep-2000
C:Accession: I47105
R:Rogers, G.R.; Hickford, J.G.; Bickerstaffe, R.
Anim. Genet. 25, 407-415, 1994
A:Title: Polymorphism in two genes for B2 high sulfur proteins of wool.
A:Reference number: I47105; MUID:95209146; PMID:7695121
A:Accession: I47105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-182 <ROG>
A:Cross-references: GB:L33885; NID:g499871; PIDN:AA801444.1; PID:g940355
C:Superfamily: Keratin high-sulfur matrix protein IIIA

Query Match 29.1%; Score 432; DB 2; Length 182;
Best Local Similarity 40.7%; Pred. No. 6.5e-24;
Matches 74; Conservative 25; Mismatches 57; Indels 26; Gaps 5

QY      43  SCGISCCCR-PSCCISSCCKPSCCLTTCRTTCRPSCCISSCCRPSCCISSCCKPSCCR 101
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2  ACCSTSFSGFPGICSTGGTGGSPCCQTCQTSQCPQTSQTSQCPQTSQTSQCPQISIO 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      102  TTCRPSCCISSCCRPSCCISSCCKPSC-----CRITTCRPSCCIS 143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62  TSCCOPTSIQTSCCOPTCLQTSGGCTGGCIGGSIQGVQSSGAVSSRTWCRPDCRVEG 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      144  CRRPSCCISSCCKPSCC-----QITTCRPSCCISSCYRPO-CCOPSCRRPACCISSCCCHP 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122  TSLPPCCVVCSTPPSCCQLYYAQASCCRPSPYCGSGCCRPACCCOPTCTEPV-CEPSCCEP 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      198  SC 199
      : :
Db      181  TC 182

RESULT 11
S18946
ultra high-sulfur keratin 1 - human
N:Alternate names: UHS keratin; ultra high-sulfur matrix protein
C:Species: Homo sapiens (man)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: S18946; B36686
R:Drabant, B.; Doenecke, D.
submitted to the EMBL Data Library, December 1991
A:Description: Nucleotide sequence of a human high-sulphur keratin cDNA.
A:Reference number: S18946
A:Accession: S18946
A:Molecule type: mRNA
A:Residues: 1-169 <DRA>
A:Cross-references: EMBL:X63755; NID:g32471; PIDN:CAA5283.1; PID:g32472
R:MacKinnon, P.J.; Powell, B.C.; Rogers, G.E.
J. Cell Biol. 111, 2587-2600, 1990
A:Title: Structure and expression of genes for a class of cysteine-rich protein
A:Reference number: A36686; MUID:91115951; PMID:1703541
A:Accession: B36686
A:Molecule type: DNA
A:Residues: 1-39; Y', 41-169 <MAC>
A:Cross-references: GB:X55293; NID:g34078; PIDN:CAA39005.1; PID:g34079
C:Genetics:

```

[illegible]

[illegible]

RESULT 13

high-sulfur wool matrix protein B2A - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 15-Sep-2000
C:Accession: I47106
R:Rogers, G.R.; Hickford, J.G.; Bickerstaffe, R.
Anim. Genet. 25, 407-415, 1994
A:Title: polymorphism in two genes for B2 high sulfur proteins of wool.
A:Reference number: I47105; MUID:95209146; PMID:7695121
A:Accession: I47106
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-172 <R0G>
A:Cross-references: GB:I33886; NID:g499872; PIDN:AA801445.1; PID:g940356
C:Superfamily: keratin high-sulfur matrix protein IIIA

	Query Match	27.1%	Score 402;	DB 2:	Length 172;
	Best local Similarity	42.5%	Pred. No.	7.7e-22;	
Matches	74;	Conservative	21;	Mismatches	59; Indels 20; Gaps 6;
Qy	43	SCCISCCR-PSCCISCCRPSCCLTTCRTTCRPPSCCISCCCRPSCCISCCCKPSCCR	101	:	:
Dd	2	ACCTSECGFICPTGTCGSSPCQTCCOPTSIOITSCOPTFIOTSCCOPTFIQ	61	:	:
Qy	102	TTCRCPPCCISCCCRPSCCI-----SSCKKPSCCRTTCRPPSCCISCCCRPSCCI	151	:	:
Dd	62	TSCOPTCLQTSGETCGGIGSGYGVGVSSGAVSS--RTRWCRPCDRVEGTSLPCCV	119	:	:
Qy	152	SCKCPSC-----QTTCCRPSCCISCYRPQC-QPSCCRPACCISCCCHPC	199	:	:
Ib	120	VSTSPSCCOLLYAQAASCRPSYCGOCCRPACCFOTCTIPV-CESCSCEPT	172	:	:

RESULT 14

147107
high-sulfur wool matrix protein B2A - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 15-Sep-2000
C:Accession: 147107
R:Rogers, G.R.; Hickford, J.G.; Bickerstaffe, R.
Anim. Genet. 25, 407-415, 1994
A:Title: Polymorphism in two genes for B2 high sulfur proteins of wool.
A:Reference number: 147105; MUID:95209146; PMID:7695121
A:Accession: 147107
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-162 <ROG>
A:Cross-references: GB:I33887; NID:g499873; PIDN:AAB01446.1; PID:g940357
C:Superfamily: keratin high-sulfur matrix protein IIIA

	Query Match	25.9%	Score 384.5	DB 2	Length 162
Best Local Similarity	37.1%	pred. No. 1.2e-20			
Matches	69	Conservative 20	Mismatches 48	Indels 49	Gaps 6
Qy	28	SCCOTTCCR-TTCRCRSCCISCCRCRSCCISCCCKPSCCLTTCRTTCCRCSCCISCCR	86		
Db	2	ACGTSFCGGPICSTGTGCTGSSPCQTCOTSCQCP-----TSIOTSCQPISTOTSCCO	56		

```
Db
2 ACCSTSFCGFFPICSTGGTCGSSSPCOPTCCOTSCCOP-----TSIOTSCCOPISIOTSCCO 56
```

Search completed: June 11, 2003, 11:04:28
Job time : 19 secs

Qy	87	PSCISGCCKPSK-----	-----	CRITCCRPSCCISSCCRPSCCISNCKKPS	128
		: :	:	:	
Db	57	PTCLQTSCECTGCGIGG	IGYQGVSSGAVSSRT	RWCPRDCRVEGTS	LPCCCVVSNCTPPS
		: :	:	:	:
Qy	129	CC-----	RTTCCRPSCCISSCCRPSCCISSCCKPSCCQTT	CCRPSCCISSCYRHPQCTQPS	183
		: :	:	:	:
Db	117	CCQLVYAQASCCRP	SYCGQSCCRPAC-----	CCQPTCTIEP-----	LCPPS
Qy	184	CCRPAC	189		
		:			
Db	157	CCEPTC	162		
		:			

RESULT 15

T30136
hypothetical protein C14C11.8 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30136
R:Du, Z.; Gattung, S.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of *C. elegans* cosmid C14C11.
A:Reference number: 420742
A:Accession: T30136
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-654 <DUZ>
A:Cross-references: EMBL:U53141; PIDN:AAA96110.1; GSPDB:GNO0023; CESP:C14C11.8
A:Experimental source: strain Bristol N2; clone C14C11
C:Genetics:
A:Gene: CESP:C14C11.8
A:Map position: 5
A:Introns: 50/3; 98/3; 155/3; 182/3; 229/3; 318/3; 365/3; 522/3; 543/3; 586/3

Query Match	25.3%	Score 375;	DB 2;	Length 654;
Best Local Similarity	26.4%;	Pred. NO. 1.1e-19;		
Matches 114;	Conservative	33;	Mismatches 73;	Indels 212;
				Gaps 29;

```

QY 3 SSC-CG5VC5IQSCG-----QGLQDESC-----CRPSCCQTTTC----- 34
DB 132 SSCGCG5LSYQVSCNOCUOSQOPQIIVVQOPQOQOQSSQCMPCAPCLPSCVQSSCAPACQPM 191
QY 35 CRTTC-----CRPSC-----CISCCRPSC-----CISSC----- 59
DB 192 CSSSCVQOQOQOQIIVVQOPQOQCASSCMPSCQPSCVQOACAPACQPMCCSSSCVQOQOQOQ 251
QY 60 -----CRPSC---CLTTTC-----CR-----TTTCRPSC 79
DB 252 IVVQOPQOQOQCAPSCMPACLLSCTNGCSGNCNNQOPVVVVQOQOQOQOQTSSCPSC 311
QY 80 ---CISCCRPSC---CISSC-----KPSCRTTCCRPSC---CISSCCRPSC 119
DB 312 QPSCVQAACAPACQPMCCSSCVCFOQOQIIVVQOPQOQCTSSCPACQSSCVQOQACAPAC 371
QY 120 -----CISSC---CKPSC---RTTC---CRPSCCIS 143
DB 372 QPKCSSOCVEQOQAIIVVQOPPTSSNNCASSCPQCTPOCVQOQTCAACACQPS-CQSS 430
QY 144 CCRPSCCLISSC---CKPSCCQTTCCRPSCCI-----SSCYRPOCCQRPSC-----C 185
DB 431 CSSNAOCVQACLPSCSSCSVQOQ--QPSVVVQODSQSSC--PSACQPSCSOOCLOAONLC 486
QY 186 RPAC---CISSCCHPSCVCSRCRPFSCPTTC-----CRT 217
DB 487 QSACQPSQSSCSQNTQCVQAC-IP-SCQQTCCGOQAQPVIVVQOPQOQNCVQACQPOCQO 544
QY 218 TCFHPICGSSC: 229
DB 545 TCGSNVOCVSAC 556

```

b6
487 QACQPSCQSSCGSN'I

Qy 218 TCFHPICGSSC 229

Db 545 TCGSNVQCVŠAC 556

Search completed: June 11, 2000
Job time : 19 secs

19

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 15:10:00 ; Search time 224 Seconds
(without alignments)
7212.667 Million cell updates/sec

Title: US-09-874-062-2
Perfect score: 1101
Sequence: 1 ctgggaaccaccagagacc.....caaaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	1101	10	US-09-874-062-2
2	198	18.0	461	10	US-09-864-761-10890
3	187	17.0	2108	10	US-09-962-832-225
4	167.4	15.2	8705	9	US-10-291-230-14
5	165.4	15.0	489	10	US-09-864-761-27528
6	160.2	14.6	478	10	US-09-864-761-6314
7	146.6	13.3	403	10	US-09-864-761-22764
8	142	12.9	11872	9	US-10-092-154-1390
9	142	12.9	11872	10	US-09-764-847-1390
10	137.4	12.5	463	10	US-09-864-761-5295
11	136.6	12.4	1974	10	US-09-864-761-4747
12	132.2	12.0	464	10	US-09-864-761-6005
13	129	11.7	1101	10	US-09-874-062-2
14	126.2	11.5	696	10	US-09-864-761-21486
15	122.6	11.1	364	10	US-09-864-761-22070
16	119.2	10.8	256	10	US-09-783-590-3765
17	118.2	10.7	422	9	US-09-854-133-337
18	118.2	10.7	422	10	US-09-738-973-337
19	116.8	10.6	442	10	US-09-864-761-11569

ALIGNMENTS

RESULT 1
US-09-874-062-2
; Sequence 2, Application US/09874062
; Patent No. US20020081607A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Four Disulfide Core Domain-Containing (FDCD) Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT033P1
; CURRENT APPLICATION NUMBER: US/09/874,062
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32462
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/168,229
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-874-062-2

Query Match	100.0%	Score 1101	DB 10	Length 1101
Best Local Similarity	100.0%	Pred. No. 2.8e-260		
Matches 1101	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	CTGGGAACCCACCCAGAACCTCCACCCCTCTCAGCGCATGGTCAGCTCCTGCTGTGGCTCT 60		
Db	1	CTGGGAACCCACCCAGAACCTCCACCCCTCTCAGCGCATGGTCAGCTCCTGCTGTGGCTCT 60		
Qy	61	GTCTGCTCTGACGAGAGTGTGGTCAAGGTCGCGCCAGGAGAGTGTGCGGCGGAGC 120		
Db	61	GTCTGCTCTGACGAGAGTGTGGTCAAGGTCGCGCCAGGAGAGTGTGCGGCGGAGC 120		
Qy	121	TGCTGCCAGACCACTCTGTCAGGACCACTGCTGCGCCGCCAGCTGCTGCAATTCAGT 180		
Db	121	TGCTGCCAGACCACTCTGTCAGGACCACTGCTGCGCCGCCAGCTGCTGCAATTCAGT 180		
Qy	181	TGCTGCAAGCTCTGCTGTATCTCCAGTGTCTGCAAAACCCAGCTGCTGCTTACAC 240		
Db	181	TGCTGCAAGCTCTGCTGTATCTCCAGTGTCTGCAAAACCCAGCTGCTGCTTACAC 240		

Sequence 5592, Ap
Sequence 6987, Ap
Sequence 4, Appli
Sequence 510, App
Sequence 157, App
Sequence 12, Appli
Sequence 5097, Ap
Sequence 21820, A
Sequence 37, Appli
Sequence 31, Appli
Sequence 225, App
Sequence 1601, Ap
Sequence 62, Appli
Sequence 66, Appli
Sequence 6, Appli
Sequence 11, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 23031, A
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 105, App

20 114.4 10.4 573 10 US-09-864-761-5592
21 112.6 10.2 22585 9 US-09-764-891-6987
22 111 10.1 1852 10 US-09-969-852-4
23 110.2 10.0 1562 9 US-10-037-270-510
24 105.6 9.6 2313 9 US-09-854-133-157
25 105.6 9.6 2313 10 US-09-738-973-157
26 105.6 9.6 2314 9 US-09-764-868-12
27 103 9.4 1967 10 US-09-864-761-5097
28 102.8 9.3 648 10 US-09-864-761-21820
29 102.2 9.3 3015 9 US-09-373-658-37
30 100.2 9.1 3331 9 US-09-373-658-31
31 99.8 9.1 2108 10 US-09-962-832-225
32 99.2 9.0 14800 9 US-09-954-456-1601
33 98.6 9.0 1649 9 US-09-789-390-62
34 98.6 9.0 1787 9 US-09-789-390-66
35 98.6 9.0 3822 9 US-09-789-390-10
36 98.6 9.0 3999 9 US-09-789-390-6
37 97 8.8 4133 9 US-09-921-417-11
38 95.6 8.7 657 9 US-09-922-469-3
39 95.6 8.7 657 10 US-09-922-480-3
40 95.6 8.7 657 10 US-09-923-236-3
41 93 8.4 256 10 US-09-864-761-23031
42 92.4 8.4 1008 9 US-09-922-469-1
43 92.4 8.4 1008 10 US-09-922-480-1
44 92.4 8.4 1008 10 US-09-923-236-1
45 92.4 8.4 1325 9 US-09-746-783-105

QY	241	TGCTGCAGGACACCTGCTGCGCCGCCAGCTGCTGCAATTCAGTTGCTGCAGGCTTCC	300
Db	241	TGCTGCAGGACACCTGCTGCGCCGCCAGCTGCTGCAATTCAGTTGCTGCAGGCTTCC	300
QY	301	TGCTGTATCTCAGCTGCTGCAAAACCCAGCTGCTGCAGGACACCTGCTGCGCCGCCAGC	360
Db	301	TGCTGTATCTCAGCTGCTGCAAAACCCAGCTGCTGCAGGACACCTGCTGCGCCGCCAGC	360
QY	361	TGCTGCATTTCCAGTTGCTGCAGGCTTCCCTGCTATCTCAGCTGCTGCAAAACCCAGC	420
Db	361	TGCTGCATTTCCAGTTGCTGCAGGCTTCCCTGCTATCTCAGCTGCTGCAAAACCCAGC	420
QY	421	TGCTGCAGGACACCTGCTGCGCCGCCAGCTGCTGCAATTTCTAGTTGCTGCAGGCTTCC	480
Db	421	TGCTGCAGGACACCTGCTGCGCCGCCAGCTGCTGCAATTTCTAGTTGCTGCAGGCTTCC	480
QY	481	TGCTGTATCTAGCTGTTGCAAAACCCAGCTGCTGCCAGACACCTGCTGCGCCGCCAGC	540
Db	481	TGCTGTATCTAGCTGTTGCAAAACCCAGCTGCTGCCAGACACCTGCTGCGCCGCCAGC	540
QY	541	TGCTGTATCTCAGCTGCTACAGGCCCCAGTGTGCGCCAGCTGCTGCTGCGCCGCCGCT	600
Db	541	TGCTGTATCTCAGCTGCTACAGGCCCCAGTGTGCGCCAGCTGCTGCTGCGCCGCCGCT	600
QY	601	TGCTGCATTTCTAGTTGCTGTCATCCAGCTGCTGTGTGTCAGCTGCGCGCTGCTTTC	660
Db	601	TGCTGCATTTCTAGTTGCTGTCATCCAGCTGCTGTGTGTCAGCTGCGCGCTGCTTTC	660
QY	661	AGCTGCCGACACCTGCTGTAGAACACCTGCTGCCACCCAGCTGCTGCGGCGAGTTCT	720
Db	661	AGCTGCCGACACCTGCTGTAGAACACCTGCTGCCACCCAGCTGCTGCGGCGAGTTCT	720
QY	721	TGCTGCTGAGTGAAGCTGCTGGATTGCTGCACCTTCTGCTCTCAACCTTCAGTTCA	780
Db	721	TGCTGCTGAGTGAAGCTGCTGGATTGCTGCACCTTCTGCTCTCAACCTTCAGTTCA	780
QY	781	GCACAGATATCTATTCAGAGAACATGTGACTTCTGATGTCGTGAAACAGAGCATG	840
Db	781	GCACAGATATCTATTCAGAGAACATGTGACTTCTGATGTCGTGAAACAGAGCATG	840
QY	841	GACTGATTTGGAATAATTTTATAGTATGATATCTCTCTTTTATAGAAATTTTATCTTA	900
Db	841	GACTGATTTGGAATAATTTTATAGTATGATATCTCTCTTTTATAGAAATTTTATCTTA	900
QY	901	TTGAATCTCAATTTACAGTCAATTCACATGATGTTTACATGTTTACATTTCTTAAT	960
Db	901	TTGAATCTCAATTTACAGTCAATTCACATGATGTTTACATGTTTACATTTCTTAAT	960
QY	961	CAATATACATAATCTTCAATGGTATCTCTTCTAGATGTTTCTTCTTAATGTTTCTGTG	1020
Db	961	CAATATACATAATCTTCAATGGTATCTCTTCTAGATGTTTCTTCTTAATGTTTCTGTG	1020
QY	1021	GTATCAATTTTCTGTTGGAATTTGTTGATGTTTCTCAATTAATTAATTCATAGTTCA	1080
Db	1021	GTATCAATTTTCTGTTGGAATTTGTTGATGTTTCTCAATTAATTAATTCATAGTTCA	1080
QY	1081	GCAAAAAAAAAAAAAAAAAA 1101	
Db	1081	GCAAAAAAAAAAAAAAAAAA 1101	

RESULT 2
US-09-864-761-10890
; Sequence 10890, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 10890
LENGTH: 461
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006070.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
US-09-864-761-10890

Query Match		18.0%;	Score 198;	DB 10;	Length 461;
Best Local Similarity		71.2%;	Pred. No. 8 5e-39;		
Matches 306;		Conservative 0;	Mismatches 115;	Indels 9;	Gaps 3;
QY	2	TGGGAACCCACCTCAGAACCTCCACCTCTGACGCCATGCTCAGCTGCTGCTGCTGCTG	61		
Db	41	TGGAAACCTTACCTAGAACCTCCACCTCTGACACCATGTTAACTCTTCTGCTGCTG	100		
QY	62	TCTGCTCTCACCAGAGCTGCTGCTCAAGGTCCTGGCCAGGAGAGCTGCTGCTGCTGCTG	121		
Db	101	TCTGCTCTGACCAGGCTGCTGATCAAGGCTCTGCAAGAGACCTGCTGCTGCTGCTGCTG	160		
QY	122	GCTGCCAGACACCTGCTGCTCAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	181		
Db	161	GCTGCCAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	220		
QY	182	GCTGCCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	241		

[illegible]

RESULT 5
US-09-864-761-27528
: Sequence 27528, Application US/09864761
: Patent No. US2002048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
: FILE REFERENCE: Aecmica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661

```

: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 27528
: LENGTH: 489
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC006070.1
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
: OTHER INFORMATION: SWISSPROT HIT: O75690, EVALUE 6.00e-10
: OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALUE 3.00e-11
: OTHER INFORMATION: NT HIT: X73462.1, EVALUE 6.00e-15
US-09-864-761-27528

```

Query Match	15.0%;	Score 165.4;	DB 10;	Length 489;
Best Local Similarity	67.5%;	Pred. No. 8.9e-31;		
Matches 280:	Conservative	0;	Mismatches 126;	Indels 9;
				Gaps 3;

[illegible]

```

RESULT 6
US-09-864-761-6314/c
; Sequence 6314, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USE
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-x-1

```

```

: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 6314
: LENGTH: 478
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC006070.1
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.9
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
: US-09-864-761-6314

Query Match 14.68; Score 160.2; DB 10; Length 478;
Best Local Similarity 64.28; Pred. No. 1.7e-29;
Matches 319; Conservative 0; Mismatches 148; Indels 30; Gaps 4:

Qy 181 TGCTGCAGGCGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCA 240
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 474 TGCTGCAGGACCCTGCTGCAGGACCACTGCTGGAAGCCCA---CCACTGTGACCA 418

Qy 241 TGCTGCAGGACCCTGCTGCGCGCCAGCTGCTGTCATTTCCATTTGCTGAGGCGCTTCC 300
    ||| |||| ||| |||| ||| ||||| ||| ||||| ||| ||||| |||
Db 417 TGCAGCAGCACCCCTGTTGTCAGCGCTCTGCTGTGTGCGCCAGCTGCTGCGCAGCCT--- 361

Qy 301 TGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCCTGCTGCGCGCCAGC 360
    ||||| ||| ||| |||| ||| ||||| ||| ||||| ||| ||||| |||
Db 360 TGCTGCCACCAACTTGTGTGCAAAACACCTGCTGCAGGACCCTGCTGCGCAGCCCA-- 303

Qy 361 TGCTGCAATTTCCAGTTGCTGCAGGCGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGC 420
    ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

```

Db 302 -CTTGTGTGCCAGCTGCTGCCAGCCCTTCTGCTGCAGCACACCCCTGCTGCCAGCCACC 244

Qy 421 TGCTGCAGGAGCACCTGCTGCGCGCCAGCTGCTGTCATTTCTAGTTGCTGCAAGCTTTC 480

Db 243 TGCTGTGGGTCCAGCTGCTGCTGCGCAAAACCCAGCTGCTGGTCCAGCTGCTGTCAGCCTTAT 184

Qy 481 TGCATGATATCTAGCTGCTGTCATTCACCAACCCAGCTGCTGCCAGACCACTGCTGCCAGCTAGC 540

Db 183 TGTGGGTCCAGTTGCTGCTT-----CAGCGCTTCTGCTCAATCCACT 145

Qy 541 TGCATGATATCTAGCTGCTGTCATTCACAGCGCCAGCTGCTGCCAGCCCTGCTGCTGCTGCTGCT 600

Db 144 TGCATATCAAACTATCTGCTTTCAGCAGCACCACTGCTGCCAGGCTACCTGCTGCCAGCTTTC 85

Qy 601 TGCTGCATTTTCTAGTTGCTGCTGTCATTCACCAAGCTGCTGCTGCCAGCTGCTGCTGCTGCTGCT 660

Db 84 TGCTGCAGGAAACACCTTCTGCTGCCAGCCCACTGCTGCTGGTCCAGCTGCTGCCAGCTTTC 25

Qy 661 AGCTGCCCGGAGCACTG 677

Db 24 TGCCACCCCAACATGCTG 8

RESULT 7

US-09-864-761-22764/c

; Sequence 22764, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

; FILE REFERENCE: Aesomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

; SEQ ID NO 6314

; LENGTH: 478

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC006070.1

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.9

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4

; US-09-864-761-6314

;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 22764
;; LENGTH: 403
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:

;; OTHER INFORMATION: MAP TO AC006070.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALU 4.00e-07
;; OTHER INFORMATION: NT HIT: q17657286, EVALU 8.00e-20
;; OTHER INFORMATION: SWISSPROT HIT: P02438, EVALU 7.50e-02
US-09-864-761-22764

Query Match 13.3%; Score 146.6; DB 10; Length 403;
Best Local Similarity 65.0%; Pred. No. 3.2e-26;
Matches 266; Conservative 0; Mismatches 134; Indels 9; Gaps 3;

QY	175	TCCAGTGTGCGAGCCCTTCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCGCTG	234
DB	400	TCCCTGCTGCTCAGCCCTACCTGCTGCGAGGACCACTGCTGGAAGCCCA---CCACTGTG	344
QY	235	ACCAGCTGCTCAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTCCAGTTGCTGCAGG	294
DB	343	ACCAGCTGCTCAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTCCAGTTGCTGCAGG	284
QY	295	CCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAAGACCACTGCTGCGCG	354
DB	283	CCT---TGCTGCCACCACTTGTGCTCAAAACCACTGCTGCAAGACCACTGCTGCGCG	227
QY	355	CCCAGCTGCTGATTTCCAGTTGCTGCGAGGCTTCTGCTATCTCAGCTGCTGCAGAA	414
DB	226	CCCA---CCTGTGTGACCACTGCTGCGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTG	170
QY	415	CCCAGCTGCTGCGAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTCTAGTTGCTGCAGG	474
DB	169	CCCATCTGCTGTGGTCCAGCTGCTGCGCAAAACCACTGCTGCTGCTGCTGCTGCTGCTG	110
QY	475	CCTTCTGCTGTATCTAGCTGTTGCAAAACCCAGCTGCTGCGAGACCACTGCTGCGCG	534
DB	109	CCCAGCTGCTGTGCAACCACTACTGCGAGGAGAACCTGCTACCAACCCACGAGTGTCTAC	50
QY	535	CCCAGCTGCTGTATCTCCAGCTGCTACAGGCCCCAGCTGCTGCCAGCCCT	583
DB	49	CTGCTGTGCTGCTTAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1

RESULT 8
US-10-092-154-1390
;; Sequence 1390, Application US/10092154
;; Publication No. US20030054375A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009C1
;; CURRENT APPLICATION NUMBER: US/10/092,154
;; CURRENT FILING DATE: 2002-03-07
;; NUMBER OF SEQ ID NOS: 2003
;; Prior application removed - See File Wrapper or Palm
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1390
;; LENGTH: 11872
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-092-154-1390

Query Match 12.9%; Score 142; DB 9; Length 11872;
Best Local Similarity 60.8%; Pred. No. 4.1e-24;
Matches 305; Conservative 0; Mismatches 185; Indels 12; Gaps 4;

QY	222	CAGCTGCTGCTGACACCTGCTGCGAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTTC	281
DB	5348	CAGAAGCCCACTTCCATCCCTGACACCATGACCCCACTGTTGCTCCCTTCTGCTCAGCC	5407
QY	282	CAGTTGCTGCGAGGCTTCTGCTGATCTCAGCTGCTGCAAAACCCAGCTGCTGCAAGAC	341
DB	5408	TACCTGCTGCGAGGACCACTGCTGCGAGGACCACTGCTGCAAGACCACTGCTGCAAGAC	5464
QY	342	CACCTGCTGCGCGCCAGCTGCTGCAATTTCCAGTTGCTGCGAGGCTTCTGCTGATTTTC	401
DB	5465	ACCCTGCTGCTGAGCCCTTCTGCTGTTTCCAGCTGCTGCGAGGCT---TGCTGCAAGCC	5521
QY	402	CAGCTGCTGCAAAACCCAGCTGCTGCGAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTTC	461
DB	5522	AACCTGCTGCTCAAAACCACTGCTGCTGAGGACCACTGCTGCGAGGACCACTGCTGCAAGAC	5578
QY	462	TAGTTGCTGCGAGGCTTCTGCTGATCTCTAGCTGTTGCAAAACCCAGCTGCTGCAAGAC	521
DB	5579	CAGCTGCTGCGAGGCTTCTGCTGATGACACCACTGCTGCGAGGACCACTGCTGCAAGAC	5638
QY	522	CACCTGCTGCGCGCCAGCTGCTGATCTCAGCTGCTACAGGCCCCAGTGTGCAAGCC	581
DB	5639	CAGCTGCTGCTGAGGACCACTGCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5698
QY	582	CTCCTGCTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	641
DB	5699	CACAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5758
QY	642	CAGCTGCGGCTGCGCTTTTCACTGCTGCGAGGACCACTGCTGCTGCTGCTGCTGCTGCTG	701
DB	5759	CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5815
QY	702	CATCTGCTGCGGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	723
DB	5816	CACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5837

RESULT 9

US-09-764-847-1390
;; Sequence 1390, Application US/09764847
;; Patent No. US20020132767A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009
;; CURRENT APPLICATION NUMBER: US/09/764,847
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2003
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1390
;; LENGTH: 11872
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-847-1390

Query Match 12.9%; Score 142; DB 10; Length 11872;
Best Local Similarity 60.8%; Pred. No. 4.1e-24;
Matches 305; Conservative 0; Mismatches 185; Indels 12; Gaps 4;

QY	222	CAGCTGCTGCTGACACCTGCTGCGAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTTC	281
DB	5348	CAGAAGCCCACTTCCATCCCTGACACCATGACCCCACTGTTGCTCCCTTCTGCTCAGCC	5407
QY	282	CAGTTGCTGCGAGGCTTCTGCTGATCTCAGCTGCTGCAAAACCCAGCTGCTGCAAGAC	341
DB	5408	TACCTGCTGCGAGGACCACTGCTGCGAGGACCACTGCTGCAAGACCACTGCTGCAAGAC	5464


```

QY 342 CACCTGCTGCGCGCCAGCTGCTGCAATTTCCAGTTGCTGCAGCCCTTCCCTGCTGTATCTC 401
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5465 ACCCTGCTGTCAGCCCTCTCTGCTGTGTTTCCAGCTGCTGCCAGCCT---TGCTGCCACCC 5521
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 402 CAGCTGCTGCAAAACCCAGCTGCTGCAGGACCACCTGCTGCCGCCCCAGCTGCTGCAATTC 461
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5522 AACTTGTGTCAAAACACCTGCTGTAGGACCAGCTGCTGCCAGCCCA---TCTGTGTGAC 5578
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 462 TAGTTGCTGCAGGCTTCTCTGCTGTATCTCTAGCTGTTGCAAAACCCAGCTGCTGCCAGAC 521
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5579 CAGCTGCTGCAGCCCTCTCTGCTGTAGGACCAGCTGCTGCCAGCCACATGCTGTGGTC 5638
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 522 CACCTGCTGCGCGCCAGCTGCTGCTATCTCCAGCTGCTACAGCCGCCAGCTGCTGCCAGCC 581
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5639 CAGCTGCTGTCAGAGCAGCTCTCTGCTGTCACCTGTGTGTCAGAGAACCTGCTACCACCC 5698
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 582 CTCTGCTGCGCGCGCTGCTGCTATCTCTAGTTGCTGATCCAGCTGCTGTGTGTC 641
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5699 CACAAGTGTGTTGCTGCTGCTGTTGCCCTAAACACAGAGCTGTGGCTCCAACTGCTGCCAGCC 5758
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 642 CAGCTGCGCTGCGCTTTCAGCTGCGCGAGCACCTGCTGTAGAACACCTGCTTCCACCC 701
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5759 CTGCTGCC---GCCAGCCTGCTGTGAGACCACTGCTGCAGGACCACCTGTTTCCAGCC 5815
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 702 CATCTGCTGCGCGAGTTCTTTC 723
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5816 CACCTGCTGTACAGCTGCTGC 5837
    ||||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 10

```

; US-09-864-761-5295/c
; Sequence 5295, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 5295
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006070.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; US-09-864-761-5295

Query Match, 12.5%; Score 137.4; DB 10; Length 463;
Best Local Similarity .64.0%; Pred. No. 6.5e-24;
Matches 258; Conservative 0; Mismatches 136; Indels 9; Gaps 3;

QY 233 TGACACCTGCTGTCAGAGACCACTGCTGCGCGCCAGCTGCTGCATTTCCAGTTGCTGCA 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 TGACACCTGTCAGAGACCACTGCTGCGCGCCAGCTGCTGCATTTCCAGTTGCTGCA 365
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 GGCTTCTGCTGCTATCTCCAGCTGTGCAAAACCCAGCTGCTGCAGAGCACTGCTGCG 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 AGCTTACTGC---CAGCCAACTTGCCTCAAAACACCTGCTGCAGAGCACTGCTGCG 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 353 GCGCCAGCTGCTGCAATTTCCAGTTGCTGCGAGGCTTCCCTGCTGCTATCTCCAGT 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 AGCCCA---CCTGTGTGTCAGAGCTGCTGCGAGCCTTCTGCTGCAGCACACCTGCT 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 413 AGCCAGCTGCTGCAGAGCACTGCTGCGCGCCAGCTGCTGCATTTCCAGTTGCTGCA 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 AGCCCACTGCTGCTGCGGTCAGAGCTGCTGCGGCAACAGCTGCTGGGTGCGAGTGT 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 473 GGCTTCTGCTGCTATCTCTAGCTGTTGCAAAACCCAGCTGCTGCAGAGCACTGCTG 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 AGAGCAGCTCCTGTGCGACCTGTGTACTGCAGAGAAGAACTGCTACACCCAGACTGCT 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 533 GCGCCAGCTGCTGCTATCTCCAGCTGCTTACAGGCCCCAGTGCCTGCCAGCCCTGCT 592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 GCCTGCTGCTGCTCAACAGAGCTGTGATCCAGCTGCTGCGAGC---CCTGCTACT 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 593 GCGCGGCTTGTGCTGCTATTTCTAGTTGCTGTATCCAGCTGCTG 635
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 73 GCGCAGCTGCTGCTGCTGCTGCTGCTGCCAGCTTCTTTGTTG 31
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
; US-09-864-761-4747
; Sequence 4747, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04

```



```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21486
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003958.1
; OTHER INFORMATION: EXPRESSED IN HT474, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: NT HIT: L33886.1, EVALUE 9.60e-02
; OTHER INFORMATION: EST_HUMAN HIT: A1907096.1, EVALUE 1.00e-116
; OTHER INFORMATION: SWISSPROT HIT: P08131, EVALUE 9.00e-16
US-09-864-761-21486

Query Match 11.5%; Score 126.2; DB 10; Length 696;
Best Local Similarity 52.4%; Pred. No. 4.8e-21;
Matches 312; Conservative 0; Mismatches 268; Indels 15; Gaps 1;

QY 119 GCTGCTGCCAGACCACTGCTGCAGGACCACCTGCTGCCGCCGCCAGCTGCTGCATTTCCTCA 178
Db 1 GCTGCTGCCAGCCCTGTAGGCTCTCAAGCCACTTCTGCGCAACCAAGTCTGCTGTGCGCCA 60
QY 179 GTTGTGCGAGCCCTTCCTGCTGTATCTCCACGCTGCTGCAAAACCCAGCTGCTGCCTGACCA 238
Db 61 CTTCTGCGACCCCTGCTCTGCAAAATCCACCTGCTGCCAGCCAGTTCCTGTGTGAGGCCA 120
QY 239 CTTGTGCGAGGACCACTGCTGCCGCCGCCAGCTGCTGCATTTCAGTTGCTGCAGGCCCTT 298
Db 121 GCTGCTGTTACGCTGCTGCACCCCTGCCTAGTCTGCGCAACCTGCTGCTGTGTGAGCCTT 180
QY 299 CTTGCTGTATCTCAGCTGCTGCGAAACCACTGCTGCAGGACCACCTGCTGCGCCGCCCA 358
Db 181 CTTGCTGTACGCCGGTGTGCCCGACACCTACCTGCTGTGACCAAGTAGCTGCCAGGCTG 240

```

```

359 GCTGCTGCATTTCAGCTTGTGTCAGGCTTCTCTGCTGTATCTCCAGCTGCTGCAGAAACCA 418
Db 241 TCTGCTGTGACCCAG-----CCTTCTGAGCCAAAGTTGCTCAGAGTCTA 285
QY 419 GCTGCTGCAGGACCACTGCTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478
Db 286 GCATCTGCCAGCCAGTACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
QY 479 CTTGCTGTATCTTAGCTGTTGCAAAACCCAGCTGCTGCCAGACCACCTGCTGCTGCTGCTG 538
Db 346 TCTGCTGTGTTTACAGCTGCTGCGAGCCACCTTCTGTCCCGAGCATTGTCGCAAAAGAGCTT 405
QY 539 GCTGCTGTATCTCCAGCTGTACAGGCCCCAGCTGCTGCCAGCCCTCTGCTGCTGCTGCTG 598
Db 406 CTTGCTGTGCTCCAGTATCTGCCAACCCATCTCTGCTGAGCCAGCCCTGCTGCTGCTGCTG 465
QY 599 CTTGCTGTATCTTAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658
Db 466 CTTGCTGTGCTGCCAGTCCATGCCAACCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
QY 659 TCAGCTGCCGCGACCACTGCTGTAGAACACCTGCTGCCAGCCATCTGCTGCTGCTGCTG 713
Db 526 TCTGCCCTGAGCCAGTTCCTGCGCCATCTACCTCTCTGCCGACCTCTTTCTCTGCTGAG 580

```

RESULT 15

```

US-09-864-761-22070/G
; Sequence 22070, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30

```

: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
: SEQ ID NO 22070
: LENGTH: 364
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC006070.1
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
: OTHER INFORMATION: SWISSPROT HIT: P45931, EVALUE 1.50e+00
: OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALUE 2.00e-05
: OTHER INFORMATION: NT HIT: M27685.1, EVALUE 2.00e-17
US-09-864-761-22070

Query Match		11.1%	Score 122.6;	DB 10;	Length 364;
Best Local Similarity		64.0%	Pred. No. 2.4e-20;		
Matches 236;		Conservative	0;	Mismatches 124;	Indels 9; Gaps 3;
QY	118 AGCTGCTGCCAGACCACTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCATTTC	177			
Db	118 AGCTGCTGCCAGACCACTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCATTTC	177			
QY	364 AGAAGCCACCTCCACCCCTGACACATGACCCACTGTTGCCCTGCTGCAGCCT	305			
Db	364 AGAAGCCACCTCCACCCCTGACACATGACCCACTGTTGCCCTGCTGCAGCCT	305			
QY	178 AGTTGCTGCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCCTGACC	237			
Db	178 AGTTGCTGCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCCTGACC	237			
QY	304 ACCTGCTGCAGGACCACTGCTGCAGGACTACCTGTGGCAGCCCA--CCATTGTGACC	248			
Db	304 ACCTGCTGCAGGACCACTGCTGCAGGACTACCTGTGGCAGCCCA--CCATTGTGACC	248			
QY	238 ACCTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCAGTTTCCAGTTGCTGCAGGCT	297			
Db	238 ACCTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCAGTTTCCAGTTGCTGCAGGCT	297			
QY	247 ACCTGCAGCAGCAGCAGCTGCTGCCAGCCCTCTGCTGTGTGTCAGCTGCTGCCAGCCT	188			
Db	247 ACCTGCAGCAGCAGCAGCTGCTGCCAGCCCTCTGCTGTGTGTCAGCTGCTGCCAGCCT	188			
QY	298 TCCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCCGCC	357			
Db	298 TCCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCCGCC	357			
QY	187 TACTGC---CACCCACTTCTGTCAAACACCTGCTGCAGGACCACTGCTGCCAGCCC	131			
Db	187 TACTGC---CACCCACTTCTGTCAAACACCTGCTGCAGGACCACTGCTGCCAGCCC	131			
QY	358 AGCTGCTGCATTTCACATTTGCTGCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACC	417			
Db	358 AGCTGCTGCATTTCACATTTGCTGCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACC	417			
QY	130 A---CCTGTGTGACCACTGCTGCCAGCCTTCTGCTGCAGCAGCAGCCTGCTGCCAGCCC	74			
Db	130 A---CCTGTGTGACCACTGCTGCCAGCCTTCTGCTGCAGCAGCAGCCTGCTGCCAGCCC	74			
QY	418 AGCTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCAGTTTCTAGTTGCTGCAGGCT	477			
Db	418 AGCTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCAGTTTCTAGTTGCTGCAGGCT	477			
QY	73 ATCTGCTGTGGTCCAGCTGCTGTGGCCAAACAGCTGTGGTCCAGCTGTGGCCAGAGC	14			
Db	73 ATCTGCTGTGGTCCAGCTGCTGTGGCCAAACAGCTGTGGTCCAGCTGTGGCCAGAGC	14			
QY	478 TCCTGCTGT 486				
Db	478 TCCTGCTGT 486				
QY	13 AGCTCCTGT 5				
Db	13 AGCTCCTGT 5				

Search completed: June 18, 2003, 17:55:55
Job time : 231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 10:15:25 ; Search time 70 seconds
(without alignments)
4823.587 Million cell updates/sec

Title: US-09-874-062-2

Perfect score: 1101

Sequence: 1 ctgggaaccacccagacc.....caaaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCRUS.COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	252	22.9	3489	2	US-08-728-323A-1
C 2	252	22.9	3489	4	US-09-298-568-1
C 3	252	22.9	32207	2	US-08-770-379-20
C 4	252	22.9	32207	4	US-08-757-669A-20
C 5	252	22.9	32207	4	US-09-230-371A-20
C 6	167.4	15.2	1926	4	US-09-249-585A-2
C 7	167.4	15.2	2580	3	US-09-050-863-2
C 8	167.4	15.2	2580	4	US-09-359-081-2
C 9	167.4	15.2	5452	2	US-09-130-114-1
C 10	167.4	15.2	9600	4	US-08-910-647-1
C 11	167.4	15.2	9600	4	US-09-620-925-1
C 12	167.4	15.2	10596	1	US-07-884-811-15
C 13	167.4	15.2	10596	1	US-07-885-971-15
C 14	167.4	15.2	10596	1	US-08-087-783A-15
C 15	167.4	15.2	10596	1	US-08-194-088B-15
C 16	167.4	15.2	10596	1	US-08-194-087-15
C 17	167.4	15.2	10596	5	PCT-US93-04648-15
C 18	151	13.7	2824	2	US-09-010-928B-3
C 19	139.4	12.7	2830	2	US-09-010-928B-1
C 20	134.6	12.2	3489	2	US-08-728-323A-1
C 21	134.6	12.2	3489	4	US-09-298-568-1
C 22	134.6	12.2	32207	2	US-08-770-379-20
C 23	134.6	12.2	32207	4	US-08-757-669A-20
C 24	134.6	12.2	32207	4	US-09-230-371A-20
C 25	128.6	11.7	2793	1	US-08-209-747-1
C 26	128.6	11.7	2793	1	US-08-458-298-1
C 27	126.6	11.5	9551	1	US-08-056-200-93

C 28	126.6	11.5	9551	2	US-08-800-644-93	Sequence 93, Appl Patent No. 5273901
C 29	122.4	11.1	543	6	5273901-6	Sequence 1, Appl Patent No. 5273901
C 30	112.8	10.2	2338	1	US-08-425-069-1	Sequence 1, Appl Patent No. 5482709
C 31	112.8	10.2	2338	2	US-08-317-844B-1	Sequence 1, Appl Patent No. 5482709
C 32	112.4	10.2	533	6	5482709-5	Sequence 1, Appl Patent No. 5482709
C 33	111.2	10.1	2830	2	US-09-010-928B-1	Sequence 1, Appl Patent No. 5482709
C 34	106.2	9.6	1995	1	US-08-425-069-3	Sequence 3, Appl Patent No. 5273901
C 35	106.2	9.6	1995	2	US-08-317-844B-3	Sequence 3, Appl Patent No. 5273901
C 36	105.6	9.6	2313	4	US-09-370-838-157	Sequence 157, Appl Patent No. 5273901
C 37	104.2	9.5	543	6	5273901-6	Sequence 3, Appl Patent No. 5273901
C 38	101.4	9.2	397	3	US-09-253-691-3	Sequence 3, Appl Patent No. 5273901
C 39	100.2	9.1	2214	3	US-08-864-038A-1	Sequence 2, Appl Patent No. 5273901
C 40	100.2	9.1	3331	3	US-08-864-038A-2	Sequence 4, Appl Patent No. 5231168
C 41	100.2	9.1	3331	3	US-08-864-038A-4	Sequence 4, Appl Patent No. 5231168
C 42	99	9.0	3095	6	5231168-1	Sequence 1, Appl Patent No. 5231168
C 43	95.6	8.7	657	4	US-09-527-345-3	Sequence 1, Appl Patent No. 5231168
C 44	93.4	8.5	2793	1	US-08-209-747-1	Sequence 1, Appl Patent No. 5231168
C 45	93.4	8.5	2793	1	US-08-458-298-1	Sequence 1, Appl Patent No. 5231168

ALIGNMENTS

RESULT 1

US-08-728-323A-1/c

; Sequence 1, Application US/08728323A

; Patent No. 5948676

; GENERAL INFORMATION:

; APPLICANT: Chang, Yuan

; APPLICANT: Bohenzky, Roy A.

; APPLICANT: Russo, James J.

; APPLICANT: Edelman, Isidore S.

; APPLICANT: Moore, Patrick S.

; TITLE OF INVENTION: Immediate Early Protein From Kaposi's

; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA

; TITLE OF INVENTION: Encoding Same And Uses Thereof

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/728,323A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-278-0400

; TELEFAX: 212-391-0525

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3489 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: Single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CUS

; LOCATION: 1..3489

US-08-728-323A-1

Query Match 22.9% Score 252; DB 2; Length 3489;

```
Best Local Similarity 59.3%; Pred. No. 1.7e-42;
Matches 429; Conservative 0; Mismatches 295; Indels 0; Gaps 0;
QY 42 CAGTCTCTGCTGGCTCTGCTCTGTGACACAGAGCTGTGGTCAAGGTCCTGCGCCAGGA 101
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2136 CTGCTCATCTGCTGCTCTGCTCATCTGCTGCTGCTCATCTGCTGCTCATC 2077
QY 102 GAGCTGTGCGCGCCAGCTGCTGCCAGACACCTGCTGCGAGGACCACTGCTGCGCGCC 161
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2076 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 2017
QY 162 CAGTGTGCTGATTCACGCTGCTGCGAGGCTTCCTGCTGATCTCCAGCTGCTGCAACCC 221
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2016 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1957
QY 222 CAGTGTGCTGACACCTGCTGCCAGGACCACTGCTGCGCGCCAGCTGCTGCTCATCTC 281
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1956 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1897
QY 282 CAGTGTGCTGACACCTGCTGCCAGGACCACTGCTGCGCGCCAGCTGCTGCTCATCTC 341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1896 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1837
QY 342 CAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 401
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1836 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTCATC 1777
QY 402 CAGTGTGCTGACACCTGCTGCCAGGACCACTGCTGCGCGCCAGCTGCTGCTCATCTC 461
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1776 CCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1717
QY 462 TAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 521
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 CCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1657
QY 522 CAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 581
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1656 CCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1597
QY 582 CTCTGCTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 641
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1596 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1537
QY 642 CAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 701
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1536 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1477
QY 702 CATCTGCTGCGCGAGTTCCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATTCCTG 761
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1476 CTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1417
QY 762 CTCT 765
  |||
Db 1416 CTGT 1413
```

RESULT 2

```
US-09-298-568-1/c
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kave, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1
```

Query Match 22.9%; Score 252; DB 4; Length 3489;

Best Local Similarity 59.3%; Pred. No. 1.7e-42;

Matches 429; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

```
QY 42 CAGTCTCTGCTGGCTCTGCTCTGTGACACAGAGCTGTGGTCAAGGTCCTGCGCCAGGA 101
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2136 CTGCTCATCTGCTGCTCTGCTCATCTGCTGCTGCTCATCTGCTGCTCATC 2077
QY 102 GAGCTGTGCGCGCCAGCTGCTGCCAGACACCTGCTGCGAGGACCACTGCTGCGCGCC 161
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2076 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 2017
QY 162 CAGTGTGCTGATTCACGCTGCTGCGAGGCTTCCTGCTGATCTCCAGCTGCTGCAACCC 221
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2016 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1957
QY 222 CAGTGTGCTGACACCTGCTGCCAGGACCACTGCTGCGCGCCAGCTGCTGCTCATCTC 281
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1956 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1897
QY 282 CAGTGTGCTGACACCTGCTGCCAGGACCACTGCTGCGCGCCAGCTGCTGCTCATCTC 341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1896 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1837
QY 342 CAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 401
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1836 CTGCTGTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTCATC 1777
QY 402 CAGTGTGCTGACACCTGCTGCCAGGACCACTGCTGCGCGCCAGCTGCTGCTCATCTC 461
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1776 CCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1717
QY 462 TAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 521
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 CCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1657
QY 522 CAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 581
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1656 CCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1597
QY 582 CTCTGCTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 641
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1596 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1537
QY 642 CAGTGTGCTGCGCGCCAGCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATCTC 701
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1536 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1477
QY 702 CATCTGCTGCGCGAGTTCCTGCTGATTCAGTGTGCTGCGAGGCTTCCTGCTGATTCCTG 761
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1476 CTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1417
QY 762 CTCT 765
  |||
Db 1416 CTGT 1413
```

RESULT 3

```
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohlenky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
```


[illegible]

```

RESULT 5
US-09-230-371A-20
: Sequence 20, Application US/09230371A
: Patent No. 6348586
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Bohenzky, Roy A
: APPLICANT: Russo, James J
: APPLICANT: Edelman, Isidore S
: APPLICANT: Moore, Patrick S
: TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
: FILE REFERENCE: 45185-G-PCT-US
: CURRENT APPLICATION NUMBER: US/09/230,371A
: CURRENT FILING DATE: 1999-11-17
: PRIOR APPLICATION NUMBER: PCT/US97/13346
: PRIOR FILING DATE: 1997-07-22
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 20
: LENGTH: 32207
: TYPE: DNA
: ORGANISM: kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

```

Query Match	22.9%	Score 252;	DB 4;	Length 32027;
Best Local Similarity	59.3%;	Pred. No. 3.2e-42;		
Matches 429;	Conservative 0;	Mismatches 295;	Indels 0;	Gaps 0;
QY	42	CAGCTCTGCTGCTGGCTCTGCTGCTCTGACACAGCTGTGTCAAGCTCTCGGCGACAGGA	101	
DB	19861	CTGCTCATCTGCTGCTCTGCTCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTATC	19920	
QY	102	GAGCTGCTGGCGCGCAGCTGCTGCGCAGACACCTGCTGCGAGGACCACTGCTGCTGCGGCTC	161	
DB	19921	CTGCTGCTGCTCATCTGCTGCTGCTGCTATCTCTGCTGCTGCTCATCTGCTGCTGCTATC	19980	
QY	162	CAGCTGCTGCATTTCACGATTCGTCGAGGCGCTTCCTGCTGTATCTCCAGCTCTGTCGAAAGC	221	
DB	19981	CTGCTGCTGCTATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTATC	20040	
QY	222	CAGCTGCTGCCCTGACACACTGCTGCGAGGACCACTGCTGCGCGCGCCAGCTGCTGCTATTC	281	
DB	20041	CTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTATC	20100	
QY	282	CAGTTGCTGCGAGCCCTTCCTGCTGTATCTGCGAGCTGCTGCGAACCAGCACTGCTGCTGCTG	341	
DB	20101	CTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTATC	20160	
QY	342	CACCCTGCTGGCGCCCGAGCTGCTGCTATTCACGATTCGCTGCGAGGCGCTCTCTGCTATTC	401	
DB	20161	CTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTG	20220	
QY	402	CAGCTGCTGCGAACCAGCTGCTGCGAGGACCACTGCTGCGCGCGCCAGCTGCTGCTATTC	461	
DB	20221	CCGCTGCTGGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC	20280	
QY	462	TAGTTGCTGCGAGCCCTTCCTGCTGTATCTGTAGCTGTTGCGAACCAGCACTGCTGCGCAGAC	521	
DB	20281	CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	20340	
QY	522	CAGCTGCTGCGCGCCAGCTGCTGTATCTCCAGCTGCTACAGGCGCCAGCTGCTGCTGCTGCTG	581	
DB	20341	CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	20400	
QY	582	CTCTCTGCTGCGCGCGCTTGCTGCGATTTCTAGTTGCTGCTCATCCAGCTGCTGCTGCTGCTC	641	
DB	20401	CTG	20460	
QY	642	CAGCTGCGCGCTGCCCTTTGAGCTGCGCGGACCACTGCTGTAGAACAACTGCTTTCACGCC	701	
DB	20461	CTG	20520	
QY	702	CATCTGCTGCTGCGCAGTTCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	761	
DB	20521	CTG	20580	
QY	762	CTCT 765		
DB	20581	CTGT 20584		

RESULT 6

```

US 09-249-595A-2/C
: Sequence 2, Application US/09249595A
: Patent No. 6417002
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert
: TITLE OF INVENTION: METHOD FOR MAIN
: FILE REFERENCE: 0867/00905
: CURRENT APPLICATION NUMBER: US/09/2
: CURRENT FILING DATE: 1999-02-11
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO: 2
: LENGTH: 1926
: TYPE: DNA
: ORGANISM: Epstein Barr Virus

```

```

; APPLICANT: HORLICK, ROBERT
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11

```

```

; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus

```

```

: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1926)
: OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Query Match      15.2%; Score 167.4; DB 4; Length 1926;
Best Local Similarity 53.0%; Pred. No. 2e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

Qy 7 ACCACCCAGAACCTCCACCTCTGACGCCATGTCAGCTCCCTGCTGCTGCTGCTGCTG 66
Db 1014 ACCCGCCCTCCACTGCTCTGACCCCGGCTCCACCTCTGCTGCTGCTGCTGCTG 955

Qy 67 TCTGACACAGCTGTGTCTGAGCTCTGCGCAGAGAGCTGCTGCGCCCGCAGTGTGC 126
Db 954 TCTGCCCCCTCTCTGCTGCTGCGCCCTCTGCGCCCTCTGCTGCTGCTGCTGCCCC 895

Qy 127 CAGACACCTCTGCTGACGAGCACCTGCTGCGCGCCCGCAGCTGCTGATTTCCAGTTGCTGC 186
Db 894 TCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 835

Qy 187 AGGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246
Db 834 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 775

Qy 247 AGGACACCTGCTGCGCGCCCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306
Db 774 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 718

Qy 307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGCGAGACACCTGCTGCTGCTGCTGCTGCTGCTG 366
Db 717 CCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658

Qy 367 ATTTCCAGTTGCTGCGCGCCCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 426
Db 657 TCCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598

Qy 427 AGGACACCTGCTGCGCGCCCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486
Db 597 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 538

Qy 487 ATCTCTAGCTGTTGCAAAACCCAGCTGCTGCGAGACACCTGCTGCTGCTGCTGCTGCTGCTG 546
Db 537 TCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478

Qy 547 ATCTCCAGCTGCTACAGGCGCCCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 606
Db 477 TGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418

Qy 607 ATTTCTAGTTGCTGTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666
Db 417 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 358

Qy 667 CCGACACCTGCTGTAGAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 726
Db 357 TCCTGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 298

Qy 727 T 727
Db 297 T 297

RESULT 7
US-09-050-863-2/c
: Sequence 2, Application US/09050863
: Patent No. 6114111
: GENERAL INFORMATION:
: APPLICANT: Lao, Ying
: APPLICANT: Hiang, Betty
: APPLICANT: Payan, Don
: TITLE OF INVENTION: Mammalian Protein Interaction Cloning
: TITLE OF INVENTION: System

```

```

: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/050.863
: FILING DATE: 30-MAR-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REGISTRATION NUMBER: 38,304
: REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 949-8711
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2580 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
US-09-050-863-2

Query Match      15.2%; Score 167.4; DB 3; Length 2580;
Best Local Similarity 53.0%; Pred. No. 2.2e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

Qy 7 ACCACCCAGAACCTCCACCTCTGACGCCATGTCAGCTCCTGCTGCTGCTGCTGCTGCTG 66
Db 1397 ACCCGCGCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1338

Qy 67 TCTGACACAGCTGTGTGTCAGGTCCTGCGCCAGGAGAGCTGCTGCGCGCTGCTGCTGCTGCTG 126
Db 1337 TCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1278

Qy 127 CAGACACCTGCTGCGAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 186
Db 1277 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1218

Qy 187 AGGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246
Db 1217 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1159

Qy 247 AGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306
Db 1157 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1101

Qy 307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 366
Db 1100 CCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1041

Qy 367 ATTTCCAGTTGCTGCGAGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 426
Db 1040 TCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 981

Qy 427 AGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486
Db 980 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 921

Qy 487 ATCTCTAGCTGTTGCAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546
Db 920 TCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 861

```



```
Qy 727 T 727
Db 726 T 726

RESULT 11
US-09-620-925-1/c
: Sequence 1, Application US/09620925
: Patent No. 6468986
: GENERAL INFORMATION:
: APPLICANT: Zuckermann et al.
: TITLE OF INVENTION: Compositions and Methods for
: Polynucleotide Delivery
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/620,925
: FILING DATE: 21-Jul-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/910,647
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Fujica, Sharon M.
: REGISTRATION NUMBER: 38,459
: REFERENCE/DOCKET NUMBER: 1218.002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 923-2706
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9600 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1
Query Match 15.2%; Score 167.4; DB 4; Length 9600;
Rest Local Similarity 53.0%; Pred. No. 3.2e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1:

Qy 7 ACCACCCAGAACCTCCACCCCTGTCAGCCATGGTCTGCTGGCTGTCTGTGC 66
Db 1443 ACCCGGGCTCCACTGCTCTCTGACCCGGCCCTCCACCTCTGCTGCGCCCTCTGC 1384
Qy 67 TCTGACCAAGAGCTGTGGTCAAGGTCTCGGCGCAGAGAGCTGTCGCGCCCGAGCTGTGC 126
Db 1383 TCCTGCGCCCTCTCTGCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCGCC 1324
Qy 127 CAGACCACTCTCAGGACCACTCTGCGCGCCCGCCAGCTGCTGCATTCCAGTTGCTGC 186
Db 1323 TCCTGCTCTGCGCCCTCTCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTCTGCTGTC 1264
Qy 187 AGGCCTTCTGCTGTATCTCCAGCTGTGTGCAAAACCCAGCTGCTGCTGACCACTGTGTC 246
Db 1263 TGCCCTCTGCGCCCTCTCTGCTGCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCTGCGCCCTCC 1204
Qy 247 AGGACCACTGCTGCGCGCCCGCCAGCTGTGTGCAATTTCCAGTTGCTGACGCGCTTCTGTGCT 306
Db 1203 TGCCCTCTGCTGCTGCGCC---CTCCTGCGCCCTCTGCTGCTGCGCCCTCTGCTGCTGCTGCTGCT 1147

307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGAGGACCACTGCTGCGCCCGCCAGCTGCTGTC 366
1146 CCCTCTGCTGCTGCGCCCTCTGCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGTC 1087
367 ATTTCCAGTTGCTGCAAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGTC 426
1086 TCCTGCGCCCTCTGCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCTGCTGCTGTC 1027
427 AGGACCACTGCTGCGCGCCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
1026 TCCTGCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCC 967
487 ATCTCTAGCTGTTGCAAAACCCAGCTGCTGCAAGACCACTGCTGCTGCGCCCGCCAGCTGCTGTC 546
966 TCCTCTGCTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907
547 ATCTCCAGCTGCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
906 TGCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
607 ATTTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666
846 TGCGCCCTCTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
667 CCGACCACTGCTGCTGCAAAACCCAGCTGCTGCAAGACCACTGCTGCTGCGCCCGCCAGCTGCTGTC 726
786 TCCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
727 T 727
726 T 726

RESULT 12
US-07-884-811-15/c
: Sequence 15, Application US/07884811
: Patent No. 5316921
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
: TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/884,811
: FILING DATE: 19920518
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dregler, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: 755.1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-3216
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: NUCLEIC ACID
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 03:06:50 ; Search time 325 Seconds
(without alignments)
7629.077 Million cell updates/sec

Title: US-09-874-062-2

Percent score: 1101
Sequence: 1 ctgggaaccaccagacc.....caaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1101	100.0	1101	22	Human cDNA clone H
2	521.4	47.4	627	23	DNA encoding novel
3	278.6	25.3	1851	22	Human polynucleoti
4	278.6	25.3	1851	22	Human polynucleoti
5	252	22.9	3489	21	Kaposi's sarcoma-a
6	252	22.9	3489	22	Nucleotide sequenc
7	252	22.9	3489	22	Kaposi's sarcoma-a
8	252	22.9	3220	20	KSHV LUR DNA (nucl
9	252	22.9	137507	19	KSHV long unique c

ALIGNMENTS

RESULT 1
AAD06303

ID AAD06303 standard: cDNA: 1101 BP.

XX AAD06303;

AC AAD06303;

DT 10-AUG-2001 (first entry)

XX Human cDNA clone HHQC55 encoding FDCD-containing protein.

DE Human: four disulfide core domain; FDCD; immunosuppressive; cytostatic;
KW antiarthritic; antirheumatic; antiproliferative; cardiast; vasotrophic;
KW cerebroprotective; nootropic; dermatological; antifungal; virucide;
KW fungicide; ophthalmological; dermatological; antifungal; virucide;
KW antiarteriosclerotic; gene therapy; skin disorder; congenital disorder;
KW mole; freckle; haemangioma; integumentary tumour; basal cell carcinoma;
KW keratosis; melanoma; atherosclerosis; urticaria; photosensitivity;
KW eczema; autoimmune disorder; lupus erythematosus; scleroderma; keloid;
KW stria; erythema; petechia; purpura; xanthelasma; autoimmune disease;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW neoplasm; angiogenesis; nervous system disorder; Alzheimer's disease;
KW infection; ocular disorder; corneal infection; clone HHQC55; ds.

XX Homo sapiens.

OS DNA encoding novel

XX Human polynucleoti

XX Human polynucleoti

XX Kaposi's sarcoma-a

XX Nucleotide sequenc

XX Kaposi's sarcoma-a

XX KSHV LUR DNA (nucl

XX KSHV long unique c

Human hair keratin
Human polynucleoti
Human polynucleoti
Human foetal liver
Human brain expres
Human bone marrow
Probe #604 used t
Human genome-deriv
Oesophagus cancer
Epstein Barr virus
EBV tethering prot
Nucleotide sequenc
Epstein-Barr virus
Anti-sense strand
Vector pShuttle DN
Vector plasmid pCM
Nucleotide sequenc
Plasmid pCisEBON f
Nucleotide sequenc
DNA clone pCEK Cl.
Nucleotide sequenc
Human foetal liver
Human brain expres
Human bone marrow
Probe #19224 used
Human genome-deriv
Epstein Barr Virus
Human foetal liver
Probe #614 for ge
Human bone marrow
Probe #6199 for ge
Probe #7430 used t
Human genome-deriv
FLGA insert stabil

PN W0200140249-A1.
 XX 07-JUN-2001.
 XX 29-NOV-2000; 2000WO-US32462.
 PF 01-DEC-1999; 99US-0168229.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX Ruben SM, Shi Y;
 XX WPI: 2001-356153/37.
 DR P-PSDB; ABE02058.
 XX
 PT Four disulfide core domains (FDCD) containing polypeptide and its
 PT polynucleotide are used to prevent, treat or ameliorate a medical
 PT condition associated with FDCD e.g. skin disorders -
 XX
 PS Claim 1: Page 229; 235pp; English.
 XX
 CC The present sequence is human cDNA clone HWHC55 encoding four
 CC disulfide core domain (FDCD)-containing protein. FDCD sequences are
 CC used to prevent, treat or ameliorate skin disorders such as congenital
 CC disorders e.g. moles, freckles and haemangiomas as well as
 CC intermediary tumours (keratosis, basal cell carcinoma, malignant
 CC melanoma), injuries and inflammation of the skin, atherosclerosis,
 CC urticaria, eczema, photosensitivity, autoimmune disorders (lupus
 CC erythematosus, scleroderma), keloids, striae, erythema, petechiae,
 CC purpura and xanthelasma. They are also used in the diagnosis and
 CC treatment of diseases and disorders of hair and epithelium.
 CC autoimmune diseases (e.g. rheumatoid arthritis), hyperproliferative
 CC disorders (e.g. neoplasms of the breast or liver), cardiovascular
 CC disorders (e.g. cardiac arrest), cerebrovascular disorders (e.g.
 CC cerebral ischaemia, angiodysplasia), nervous system disorders (e.g.
 CC Alzheimer's disease), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infection). The FDCD-
 CC containing proteins are used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. They are used in
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition. The antibodies to FDCD protein are used in alleviating
 CC symptoms associated with the disorders and in diagnostic immunoassays
 CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
 CC The FDCD sequences are also used in gene therapy.
 XX
 SQ Sequence 1101 BP; 220 A; 343 C; 231 G; 307 T; 0 other;

Query Match 100.0%; Score 1101; DB 22; Length 1101;
 Best Local Similarity 100.0%; Pred. No. 2.7e-194;
 Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGGAACCCAGAGAGCTGCTCAAGCTCTCGGCGGAGAGAGCTGCTGCGCGCCCGCAGC 120
 DB 1 CTGGGAACCCAGAGAGCTGCTCAAGCTCTCGGCGGAGAGAGCTGCTGCGCGCCCGCAGC 60
 QY 61 GTCTGCTGTGACAGAGCTGCTGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAG 120
 DB 61 GTCTGCTGTGACAGAGCTGCTGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAG 120
 QY 121 TGTGTCAGACACCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCT 180
 DB 121 TGTGTCAGACACCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCT 180
 QY 181 TGTGTCAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCT 240
 DB 181 TGTGTCAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCT 240
 QY 241 TGTGTCAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCT 300
 DB 241 TGTGTCAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCTGCTGAGAGAGCT 300

QY 301 TGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGAGAGACACCTGCTGCGGCGCCAGC 360
 DB 301 TGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGAGAGACACCTGCTGCGGCGCCAGC 360
 QY 361 TGCTGCATTTCCAGTTGCTGAGAGCTTCCCTGCTGATCTCCAGCTGCTGCAAAACCCAGC 420
 DB 361 TGCTGCATTTCCAGTTGCTGAGAGCTTCCCTGCTGATCTCCAGCTGCTGCAAAACCCAGC 420
 QY 421 TGCTGCAGGACCACTGCTGCGCGCCAGCTGCTGAGATTTCTAGTTGCTGCAAGCTTTCC 480
 DB 421 TGCTGCAGGACCACTGCTGCGCGCCAGCTGCTGAGATTTCTAGTTGCTGCAAGCTTTCC 480
 QY 481 TGCTGTATCTTAGCTTTGCAAAACCCAGCTGCTGCCAGACACCTGCTGCGAGCTGCTGCAAG 540
 DB 481 TGCTGTATCTTAGCTTTGCAAAACCCAGCTGCTGCCAGACACCTGCTGCGAGCTGCTGCAAG 540
 QY 541 TGCTGTATCTTAGCTGCTACAGGCGCCAGCTGCTGCCAGCGCTGCTGCGGCGCCAGCT 600
 DB 541 TGCTGTATCTTAGCTGCTACAGGCGCCAGCTGCTGCCAGCGCTGCTGCGGCGCCAGCT 600
 QY 601 TGCTGCATTTCTAGCTTTGCTGTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 601 TGCTGCATTTCTAGCTTTGCTGTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 QY 661 AGCTGCCGCGACCTGCTGTAGAACACCTGCTTCCACCCCATCTGCTGCGAGCTGCTGCTGCT 720
 DB 661 AGCTGCCGCGACCTGCTGTAGAACACCTGCTTCCACCCCATCTGCTGCGAGCTGCTGCTGCT 720
 QY 721 TGCTGCTGAGTGAAGCTGCTGCTGCAATTTGCTGACACCTTCTGCTCAACCTTTATTTCA 780
 DB 721 TGCTGCTGAGTGAAGCTGCTGCTGCAATTTGCTGACACCTTCTGCTCAACCTTTATTTCA 780
 QY 781 GCACAGAGTATCTATTCAGAGAACATGCGAGCTTCTGATGCTGCGTGAAGAACAGAGCATG 840
 DB 781 GCACAGAGTATCTATTCAGAGAACATGCGAGCTTCTGATGCTGCGTGAAGAACAGAGCATG 840
 QY 841 GACTGATTTGGAATAATTTTATTAGTATGTTCTCTTTTATAGAAGTTTATTTCTTA 900
 DB 841 GACTGATTTGGAATAATTTTATTAGTATGTTCTCTTTTATAGAAGTTTATTTCTTA 900
 QY 901 TTGAATCTGAATTTACAGTCAAAATTTCCACATCACATGTTTATAGAATTTCTTTATTT 960
 DB 901 TTGAATCTGAATTTACAGTCAAAATTTCCACATCACATGTTTATAGAATTTCTTTATTT 960
 QY 961 CAATATACATAAATCTTCAAAATGGTATCTTCTAGATGTTTCTTCTTAATGTTTCTG 1020
 DB 961 CAATATACATAAATCTTCAAAATGGTATCTTCTAGATGTTTCTTCTTAATGTTTCTG 1020
 QY 1021 GTATCAATTTTCCATGCTGGAATTTGTTGATGTTCTTCCCAATAAACTTCATAGTTTCAAAA 1080
 DB 1021 GTATCAATTTTCCATGCTGGAATTTGTTGATGTTCTTCCCAATAAACTTCATAGTTTCAAAA 1080
 QY 1081 GCAG 1101
 DB 1081 GCAG 1101
 RESULT 2
 AAS93555
 ID AAS93555 standard; cDNA; 627 BP.
 XX
 AC AAS93555;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29359.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 PN W0200175067-A2.

XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-0508631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB; ABG29368.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 29359; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 627 BP; 89 A; 240 C; 151 G; 147 T; 0 other;
Query Match 47.48; Score 521.4; DB 23; Length 627;
Best Local Similarity 91.5%; Pred. No. 2.6e-87;
Matches 552; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 131 CCACCTGCTGCAGGACCACTGCTGCCGCCGCCAGCTGCTGCATTTCCAGTTGCTGCAGGC 190
DB 25 CCTGTGAGGCATGGTCAGCTCCCTGCTGTGGCTGTCTGCTGCACAGAGCTGTGGTC 84
QY 191 CTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCCACCACTGCTGCAGGA 250
DB 85 AAGTCTCGGCCAGGAGCTGCTGCCGCCGCCAGCTGCTGCCAGACCACTGCTGCAGGA 144
QY 251 CCACCTGCTGCCGCCGCCAGCTGCTGCATTTCCAGTTGCTGCAGGCCCTTCTGCTGTATCT 310
DB 145 CCACCTGCTGCCGCCGCCAGCTGCTGCATTTCCAGTTGCTGCAGGCCCTTCTGCTGTATCT 204
QY 311 CCACCTGCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCCGCCGCCAGCTGCTGCATTT 370
DB 205 CCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCCGCCGCCAGCTGCTGCATTT 264
QY 371 CCAGTTGCTGCAGGCCCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGA 430
DB 265 CCAGTTGCTGAGGCCCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGA 324
QY 431 CCACCTGCTGCCGCCGCCAGCTGCTGCATTTCTAGTTGCTGCAGGCCCTTCTGCTGTATCT 490
DB 325 CCACCTGCTGCCGCCGCCAGCTGCTGCATTTCTAGTTGCTGCAGGCCCTTCTGCTGTATCT 384

QY 491 CTAGCTGTTGCAAACCCAGCTGCTGCCAGACCACTGCTGCCGCCGCCAGCTGCTGTATCT 550
DB 385 CTAGCTGTTGCAAACCCAGCTGCTGCCAGACCACTGCTGCCGCCGCCAGCTGCTGTATCT 444
QY 551 CCAGCTGCTACAGCCGCCAGCTGCTGCCAGCCCTCCTGCTGCCGCCGCCAGCTGCTGTATCT 610
DB 445 CCAGCTGCTACAGCCGCCAGCTGCTGCCAGCCCTCCTGCTGCCGCCGCCAGCTGCTGTATCT 504
QY 611 CTAGTTGCTGTCATCCAGCTGCTGCTGCTGCCAGCTGCCCTTTTCAGTGGGCTCA 670
DB 505 CTAGTTGCTGTCATCCAGCTGCTGCTGCTGCCAGCTGCCCTGCCCTGCCCTGCCCTGCCA 564
QY 671 CCACCTGCTGTAGACAACTGCTTCCACCCCATCTGCTGCCAGCTGCTGCTGTATCT 730
DB 565 CCACCTGCTGTAGAAACCTGCTTCCACCCCATCTGCTGCCAGCTGCTGCTGTATCT 624
QY 731 TGA 733
DB 625 TGA 627
RESULT 3
AAK52135/C
ID AAK52135 standard; cDNA; 1851 BP.
XX
AC AAK52135;
XX
DT 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 680.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.
OS Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPT: 2001-476283/51.
XX P-PSDB; AAM79002.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 2374-2376; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AA80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or


```
Query Match      18.3%; Score 201.2; DB 22; Length 1851;
Best Local Similarity 59.9%; Pred. No. 4.5e-28;
Matches 417; Conservative 0; Mismatches 258; Indels 21; Gaps 4;

QY 33 CGCCATGGTCAGCTCCTGCTGTGGTCTCTGTGCTGTGACAGAGCTGTGGTCAAGGTCT 92
DB 1140 CACCACCTGTGAGCACCTGTCAGGACGACACCCCTGTTGCCAGCCCTCCTGCTGTGTGCCAG 1199
QY 93 CGGCCAGGAGAGCTGCTGCGCGCCAGCTGCTGCCAGACCACTGCTGCAGGACCACTG 152
DB 1200 CTGCTGCCAGCTTGTGCGCCACCACTGCTGTCAAAACCACTGCTGCAGGACCACTG 1259
QY 153 CTGCCGCCAGCTGCTGCAATTCAGTTGCTGCTGAGGCTTCTGCTGTATCTTCCAGCTG 212
DB 1260 CTGCCAGCCCA---CTTGTGTGCCAGCTGCTGCCAGCCCTTCTGCTGTCAGCACACCTG 1316
QY 213 CTGCAAAACAGCTGTGCTGCTGACCACTGCTGTCAGGACCACTGCTGTCGCCGCCAGCTG 272
DB 1317 CTGCCAGCCAGCTGCTGTGGGTCCAGCTGCTGTGCGCCAAACAGCTGTGGGTCCAGCTG 1376
QY 273 CTGCAATTTCCAGTT-----GCTGCGAGGCTTCTGCTGTATCTCCAGCTGCTGCAA 323
DB 1377 CTGTCAGCTATTTGTGGGTCCAGTTGCTGTGAGCCCTTGTGTCACCCGACTTGTCTATCA 1436
QY 324 ACCAGCTGCTGAGGACCACTGCTGCGCGCCAGCTGCTGTCATTTCCAGTTGCTGCGAG 383
DB 1437 AACTATCTGCTTCAGGACCACTGCTGCCAGCTACTGCTGCCAGCCCACTGCTGCGAG 1496
QY 384 GCTTCTCTGCTPATCTCCAGCTGCTGCAAAACCACTGCTGTCAGGACCACTGCTGCGCG 443
DB 1497 GAACACCTCTTGCAGCCCACTGCTGTGGGTCCAGCTGCTGCCAGCTTGTGCGCACCC 1556
QY 444 CCCCAGCTGCTGATTTCTAGTTG-----CTGCGAGGCTTCTGCTGTATCTCTAGCTG 497
DB 1557 AACATGCTGTCFAAACCAATTTGTAGATCCACCTGCTGCCAACCACTCTGTGTGACCAAGATG 1616
QY 498 TTGCAAAACCCAGCTGCTGTCAGACCACTGCTGCTGCCGCCAGCTGCTGTATCTTCCAGCTG 557
DB 1617 CTGCAGCACACCTGTTTGGCAGCAACCTGCTGTGGTCCAGCTGCTGTAGCCAAACCTG 1676
QY 558 CTACAGGCGCCAGCTGCTGCTGAGCCCTTCTGCTGCGCGCGCGCTTGTGCTGCAATTTCTAGTTG 617
DB 1677 CAATGAGTCCAGCTATTTGCTGTCG---CTTGTGCTGCCCTGCCACCTGCTGCCAGACCACTG 1733
QY 618 CTGTCATCCAGCTGCTGTGTCAGCTGCGCCTTTCAGCTGCGCCGACCACTG 677
DB 1734 CTACAGGACCACTGTTTGGCGGCCAGCTGTTGCTGCACTGCTGCTGCTGCTGCTTCTCCAGCTG 1793
QY 678 CTGTAGAACAACCTGCTTCCACCCCACTGCTGCTGCGG 713
DB 1794 CTGCCAGCCTTCTGCTGCTGCTAATCCACTTGTGCTGCGAG 1829
```

RESULT 12

```
AAK53119/c
ID AAK53119 standard; cdNA; 1851 BP.
XX AC
AC AAK53119;
XX
XX
XX 06-NOV-2001 (first entry)
DT Human polynucleotide SEQ ID NO 2648.
DE
DE
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
```

```
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Weijman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR P-PSDB; AAM79986.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
PT
XX
XX Claim 1; Page 4903; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM7823-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1851 BP; 340 A; 574 C; 558 G; 379 T; 0 other;
SQ
Query Match      18.3%; Score 201.2; DB 22; Length 1851;
Best Local Similarity 59.9%; Pred. No. 4.5e-28;
Matches 417; Conservative 0; Mismatches 258; Indels 21; Gaps 4;

QY 33 CGCCATGGTCAGCTCCTGCTGTGGTCTCTGTGCTGTGACAGAGCTGTGGTCAAGGTCT 92
DB 712 CACCACCTGTGACCACTGTCAGGACGACACCCCTGTTGCCAGCCCTCCTGCTGTGTGCCAG 653
QY 93 CGGCCAGGAGAGCTGCTGCGCGCCAGCTGCTGCCAGACCACTGCTGCCAGGACCACTG 152
DB 652 CTGCTGCCAGCCCTTGTGCGCCACCACTGCTGTCAAAACCACTGCTGCCAGGACCACTG 593
QY 153 CTGCCGCCAGCTGCTGTCATTTCCAGTTGCTGCTGCGCGCTTCTGCTGTATCTTCCAGCTG 212
DB 592 CTGCCAGCCCA---CTTGTGTGCCAGCTGCTGCCAGCCCTTCTGCTGTCAGCACCACTG 536
QY 213 CTGCAAAACCCAGCTGCTGCTGACCACTGCTGTCAGGACCACTGCTGCGCGCTTCCAGCTG 272
DB 535 CTGCCAGCCCACTGCTGCTGGGTCCAGCTGCTGTGGCCAAACCACTGCTGGTTCACCTG 476
QY 273 CTGCAATTTCCAGTT-----GCTGCGAGGCTTCTGCTGTATCTTCCAGCTGCTGCAA 323
DB 475 CTGTCAGCCCTATTTGTGGGTCCAGCTGCTGTGCACTGCTGTCACCCGACTTGTCTATCA 416
QY 324 ACCAGCTGCTGAGGACCACTGCTGCGCGCCAGCTGCTGTCATTTCCAGTTCTGCTGAG 383
DB 415 AACTATCTGCTTCAGGACCACTGCTGCCAGCTTACCTGCTGCCAGCCCACTGCTGTCAG 356
QY 384 GCCTTCTGCTGTGTATCTTCCAGCTGCTGCCAAACCACTGCTGCCAGGACCACTGCTGCTGCGG 443
```


DR WPI: 2001-483446/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX Example 4; SEQ ID NO: 5829; 650pp + Sequence Listing; English.
 PS
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX
 XX Sequence 461 BP; 74 A; 171 C; 107 G; 109 T; 0 other;
 SQ
 Query Match 18.0%; Score 198; DB 22; Length 461;
 Best Local Similarity 71.2%; Pred. No. 1.3e-27;
 Matches 306; Conservative 0; Mismatches 115; Indels 9; Gaps 3;
 QY 2 TGGAAACCCACCCAGACCTCCACCCCTCTGACGCCATGGTCAAGTCTCTGCTGGTGGCTCTG 61
 DB 41 TGGAAACCTACCTAGAACCTCCACCCCTCTGACACCATGGTTAACTCTTTGTGGCTCTG 100
 QY 62 TCTGCTCGACACAGCTGTGTCAGGCTCTCGGCCAGGAGCTGCTGCGGCCCCAGCT 121
 DB 101 TCTGCTCGACAGGCTGTGATCAAGCCCTCTGCCAAGAGACCTGCTGCGGCCCCAGCT 160
 QY 122 GCTGCCAGACACCTGCTGCGGCCCCAGCTGCTGCGGCCCCAGCTGCTGCAATTTCCAGTT 181
 DB 161 GCTGCCAGACACCTGCTGCGGCCCCAGCTGCTGTTGATCCAGCTGCTGCGGCCCCATCCT 220
 QY 182 GCTGAGGCTTCTGCTGATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACACCT 241
 DB 221 GCTCTCAGACTACCTGCTGCGGCCCCAGCTGCTGCGGCCCCAGCTGCTGCGGCCCCAGCT 280
 QY 242 GCTCAGGACACCTGCTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
 DB 281 GTTGTGACAGACCA---CTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 337
 QY 302 GCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCGGCCCCAGCTGCTGCGGCCCCAGCT 361
 DB 338 GCTGTGACAGCA---CCTGCCACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 394
 QY 362 GCTGATTTCCAGTTGCTGCGGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
 DB 395 GTTGTGACAGCA---CCTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 451
 QY 422 GCTCAGGAC 431
 DB 452 GCTGTGACAG 461
 RESULT 15
 AAK31470
 ID AAK31470 standard; DNA; 461 BP.
 XX
 AC AAK31470;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 6027.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157276-A2.
 XX
 PD 09-AUG-2001.
 XX

PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-488900/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX Example 4; SEQ ID NO: 6027; 658pp + Sequence Listing; English.
 PS
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 XX Sequence 461 BP; 74 A; 171 C; 107 G; 109 T; 0 other;
 SQ
 Query Match 18.0%; Score 198; DB 22; Length 461;
 Best Local Similarity 71.2%; Pred. No. 1.3e-27;
 Matches 306; Conservative 0; Mismatches 115; Indels 9; Gaps 3;
 QY 2 TGGAAACCTACCTAGAACCTCCACCCCTCTGACGCCATGGTCAAGTCTCTGCTGGTGGCTCTG 61
 DB 41 TGGAAACCTACCTAGAACCTCCACCCCTCTGACACCATGGTTAACTCTTTGTGGCTCTG 100
 QY 62 TCTGCTCGACACAGCTGTGTCAGGCTCTCGGCCAGGAGCTGCTGCGGCCCCAGCT 121
 DB 101 TCTGCTCGACAGGCTGTGATCAAGCCCTCTGCCAAGAGACCTGCTGCGGCCCCAGCT 160
 QY 122 GCTGCCAGACACCTGCTGCGGCCCCAGCTGCTGCGGCCCCAGCTGCTGCAATTTCCAGTT 181
 DB 161 GCTGCCAGACACCTGCTGCGGCCCCAGCTGCTGTTGATCCAGCTGCTGCGGCCCCATCCT 220
 QY 182 GCTGAGGCTTCTGCTGATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACACCT 241
 DB 221 GCTCTCAGACTACCTGCTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280
 QY 242 GCTCAGGACACCTGCTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
 DB 281 GTTGTGACAGACCA---CTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 337
 QY 302 GCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCGGCCCCAGCTGCTGCGGCCCCAGCT 361
 DB 338 GCTGTGACAGCA---CCTGCCACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 394
 QY 362 GCTGATTTCCAGTTGCTGCGGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
 DB 395 GTTGTGACAGCA---CCTGCGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 451
 QY 422 GCTCAGGAC 431
 DB 452 GCTGTGACAG 461
 Search completed: June 18, 2003, 10:20:21
 Job time : 337 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: June 11, 2003, 11:05:01 ; Search time 45 Seconds
(without alignments)
527.672 Million cell

Title: US-09-874-062-3

Perfect score: 1484

Sequence: 1 MVSSCCGVCSDQSCGQLG.....PTTCRRTCFHPICGSSCC 230

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length:	200000000
Maximum DB seq length:	200000000

Post-processing: Minimum Match 08

Loss processing: Minimum Match 0% Maximum Match 100%

Maximum match 100%
Listing first 45 summaries

```
Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*
```

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1484	100.0	230	10	US-09-874-062-3	Sequence 3, Appli
2	513	34.6	231	10	US-09-864-761-37946	Sequence 37946, A
3	489	33.0	3734	9	US-10-123-155-147	Sequence 147, App
4	479.5	32.3	1904	9	US-10-123-155-99	Sequence 99, Appl
5	479	32.3	1732	9	US-10-184-644-281	Sequence 281, App
6	479	32.3	1732	9	US-10-184-634-281	Sequence 281, App
7	466.5	31.4	134	10	US-09-864-761-39564	Sequence 39564, A
8	448	30.2	102	10	US-09-864-761-43621	Sequence 43621, A
9	443	29.9	2294	9	US-10-184-644-283	Sequence 283, App
10	443	29.9	2294	9	US-10-184-634-283	Sequence 283, App
11	439	29.6	2651	9	US-10-184-644-135	Sequence 135, App
12	439	29.6	2651	9	US-10-184-634-135	Sequence 135, App
13	438.5	29.5	3721	9	US-10-123-155-543	Sequence 543, App
14	437.5	29.5	1813	9	US-10-123-155-337	Sequence 337, App
15	437	29.4	2236	9	US-10-123-155-381	Sequence 381, App
16	436.5	29.4	702	9	US-10-123-155-193	Sequence 193, App
17	434.5	29.3	1901	9	US-10-184-644-47	Sequence 47, Appl
18	434.5	29.3	1806	9	US-10-184-634-47	Sequence 47, Appl
19	433	29.2	2806	9	US-10-123-155-201	Sequence 201, App

ALIGNMENT'S

RESULT 1

US-09-874-062-3

; Sequence 3, Applicant

; Patent No. US20020081607
CENTRAL INTELLIGENCE

```

; GENERAL INFORMATION:
; APPLICANT : Ruben et al.
; TITLE OF INVENTION : Four Disulfide Core Domain-Containing Polynucleotides, Polypeptides, and Antibodies
; TITLE OF INVENTION : Polynucleotides, Polypeptides, and Antibodies

```

	Query Match	100.0%;	Score 1484;	DB 10;	Length 230;
	Best Local Similarity	100.0%;	Pred. No. 5.1e-88;		
	Matches 230;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MVSSCCGVCSDQSGOGLQGESCCRPSCCOTTCRTTCCRPSCCISSCCRPSCCIS	SICC	60	
Dd	1	MVSSCCGVCSDQSGOGLQGESCCRPSCCOTTCRTTCCRPSCCISSCCRPSCCIS	SICC	60	
Qy	61	KPSCCLTTCRTTTCRRPSCCISSCCRPSCCISSCKKPSCCRTTTCRRPSCCISSCKHP	SICC	120	
Dd	61	KPSCCLTTCRTTTCRRPSCCISSCCRPSCCISSCKKPSCCRTTTCRRPSCCISSCR	PPSCC	120	
Qy	121	ISSCKKPSCCKTTCRRPSCCISSCCRPSCCISSCKKPSCCOTTTCRRPSCCISSCK	VPDCC	180	
Dd	121	ISSCKKPSCCKTTCRRPSCCISSCCRPSCCISSCKKPSCCOTTTCRRPSCCISSCK	VPDCC	180	
Qy	181	QPSCCRRPACCISSSCHIPPSCCVSSCRCRPFSCPTTCRTTCFHPICGGSSCC	230		
Dd	181	QPSCCRRPACCISSSCHIPPSCCVSSCRCRPFSCPTTCRTTCFHPICGGSSCC	230		

```

RESULT 2
US-09-864-761-37946
; Sequence 37946, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37946
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003958.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: P08131, EVALUE 9.00e-16
; OTHER INFORMATION: EST_HUMAN HIT: AI907096.1, EVALUE 3.00e-39
US-09-864-761-37946

```

```

Query Match      34.6%:   Score 513:   DB 10:   Length 231:
Best Local Similarity 40.6%:   Pred. No. 2,4e-26;
Matches      86:   Conservative      29:   Mismatches      77:   Indels      20:   Gaps      3:

Qy      29      CCQTTCCTTCCKRPSCCISSCCKRPSCCISSCCKRPSCCITTCCTTCCKRPSCCISSCCKRPS 88
      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      3      CCQVWGSEATSCQVLVLCVPTSCQVLCKSCCQ-      -----VCEPSCCSAVCTLPS 50

Qy      89      CCISSCCKPSCCRTTCCKRPSCCISSCCKRPSCCISSCCKRPSCCRTTCCKRPSCCISSCCKRPS 143
      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      51      SCQPVVCEPSCCOPVCTPTCTSVTSSCQAVCCDPSPCEPSCSESSICQATCVALVCEPV 110

Qy      144      CCRPSCCISSCCKRPSCCQTTTCCKRPSCCISSCCKRPSCCISSCCKRPSCCQCPQCCCKRPSCCISSCCKRPS 199
      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      111      CLREVCQVSCSEPPSPVSTCQEPSCGVSSICQPICEPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 170

Qy      200      -CVSSCKRPSCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 230
      |||      |||      |||      |||      |||      |||      |||      |||      |||      |||
Db      171      KCPVSCVCEPVPVSCPTSCRPISCSFGSSASAIC 202

RESULT 3
US-10-123-155-147
; Sequence 147: Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeFurque, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 147
; LENGTH: 3734
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-123-155-147

```

[illegible]

Qy 161 QT-----PSCCISCYRPQCC---OPSCCRPACCISSCCHPSOCVSSCRCPFS 210
Db 3122 ATATCTCCACTGCTACCTTGCTGCCCGCAGAGCC--ACCTGCGCCAAACCAAC-CACTC 3178
Qy 211 P-----TTCRTTCTCFHPICCCSSCC 230
Db 3179 AGGTAACGCCCACTAATCAGCAGGGGCC 3206

RESULT 4

US-10-123-155-99
; Sequence 99, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 99
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-99

Query Match 32.3%; Score 479.5; DB 9; Length 1904;

Best Local Similarity 30.8%; Pred. No. 1.le-23;
Matches 94; Conservative 22; Mismatches 90; Indels 99; Gaps 19;

Qy 3 SSCGSGVSDGSCG-----QGLGQESCRPSCCCTTCC-----RTTCCRP 42
Db 1397 TTTACCCCTTGCGCTAGGCTGCG--CACTCCCTTTCCCGCAGCTTTAATACTCCTG 1453
Qy 43 SCC---ISSCCRPSCCISCCKPSCLTTCGRTTCRPSCCISSC-CRPSCCISSCCRPS 98
Db 1454 GCGTGGCACCTCAACCCACCGCTACTTCCCATCC---CCGAGCGCTGTGCTGCTTCA 1510
Qy 99 CRRTCCRPSCISS-----CCRPS-----CCISSCKPSCRRTCCRPS 138
Db 1511 CCATACCCCGGCTAGACTGTAAAGCGCTAAAGCCTCGGCTGTCC---TCC-CACCATT 1567
Qy 139 C--CISSCCRPS--CCISSCKPSCCT-----TCCRPSCCIS--CY 175
Db 1568 CTGCTGCGCATGCTGCTGCC---CCTTTCTCCAAACCTATTAGGTACCGGAACA 1624
Qy 176 RPQCC-----QPSCCRPACCISSCCHPSOCVSSCRCPFSCTTTCRTTCTCFHPITCC 225
Db 1625 GAACCCCTGGGCTGAGGCGCTGCGCTGCC---CCCGGCC---CTGCCCTTGC-----CC 1674
Qy 226 GSSCC 230
Db 1675 GCCCC 1679

RESULT 5

US-10-184-644-281
; Sequence 281, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William J.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 281
; LENGTH: 1732
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-281

Query Match 32.3%; Score 479; DB 9; Length 1732;

Best Local Similarity 33.2%; Pred. No. 1.le-23;
Matches 97; Conservative 27; Mismatches 72; Indels 96; Gaps 21;

Qy 5 CCGSVCSDSGCGGLGQESCRPSCCCTTCC-----CRRTCCRP----- 42
Db 2 CCACGCG--TCCGCG-----CCTCTCCCTTCTGCTGGACCTTCCCTCGTCTCCATCTCT 55
Qy 43 ---SCCISCKRPSG-----CISCKPSCCLTTCRRTTC--CRPSCCIS--SCCRPS;C 90
Db 56 CCCTCTTTCCCGGCTCTCTTCC--ACCTTCTCTCTTCCACCTTAGAGC--TCC 111
Qy 91 ISSCKPSCRRTTCRPSCCISSCCRPSCC-----ISSCKPSCRRT--TCCRPSCCISSC 144
Db 112 CTCTCTGCGC--TCTTTCTCTGCGC--ACCGCTGCTCTCTGCGCCTTCTCCGACCCCT 167
Qy 145 CRPSC--CISCKRPSCCQ-----TT-----CRRPSCCISCKRPS 179
Db 168 CTAGCAGCAGAGC--TCTGGGGTCTGTGGTTCATCTGTGCGCCTCTGTGCTTCTGCTC 225
Qy 180 CUPSCKRP--ACCISSCKHPSCVSSCKRCPFSCTTTCRTTCTCFHPICCGSSC;C 240
Db 226 CTTTTCGCTCTGCTTCC--TCCGAGC-----TCCGCTC-----CCGAGAC 263

RESULT 6

US-10-184-634-281
; Sequence 281, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William J.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217


```

; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 281
; LENGTH: 1732
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-281

Query Match          32.3%; Score 479; DB 9; Length 1732;
Best Local Similarity 33.2%; Pred. No. 1,1e-23;
Matches 97; Conservative 27; Mismatches 77; Indels 96; Gaps 21;

UY 5 CCGSVCSQSCGGGQESCCRPSCCQTTC-----CRITCCRP----- 42
    || | || | || | || | || | || | || | || | || | || |
Db 2 CCAGCGG-TCCGCG-----CCTCTCCCTTCGCTGGACCTCCTTCGCTCTCATCTCT 55
    || | || | || | || | || | || | || | || | || | || |

QY 43 ---SCCISCCRPSC-----CISSCCKPSCCLTTCCRTTC--CRPSCCIS--SCCRPSCC 90
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 56 CCCTCCTTTCCCGGGTTCTCTTTCC--ACCTTTCTCTTCCACCTTAGACC--TCC 111
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 91 ISSCKPSCRTTCCRPSCCISSCCRPSCC-----ISSCKPSCCRT--TCCRPSCCISSC 144
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 112 CTTCCTGCCC--TCCTTTCTCTGCC--ACCGCTGCTTCTCGCCCTTCCGACCCCGCT 167
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 145 CRPSC--CISSCCKPSCC-----TT-----CCRPSCCISSCYRPQC 179
    | | | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 168 CTAGACGACGACC--TCCTGGGCTGTGGGTGTGATCTGCGCCCTCTGCTCGGTGC 225
    | | | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

QY 180 COPSCCRP-ACCISSCCHPSCVSCRPFCFSCPTTCCMTTCFHPHICCGSSCC 230
    | | | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 226 CTTTTCGCTCCCTTCC--TCCGAC-----TCCGCTC---CCGGAC 263
    | | | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 7
US-09-864-761-39564
; Sequence 39564, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43621
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006070.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALUATE 2.00e-12
; OTHER INFORMATION: SWISSPROT HIT: O75690, EVALUATE 3.00e-10
US-09-864-761-43621

Query Match 30.2%; Score 448; DB 10; Length 102;
Best Local Similarity 51.3%; Pred. No. 2e-22;
Matches 77; Conservative 11; Mismatches 14; Indels 48; Gaps 5;
QY 25 CRPSCCCTTCRTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCC 84
Db 1 CRPSCCCTTCCTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCC 29
QY 85 CRPSCCISCCCKPSCCTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCC 144
Db 30 CRPSCGVSSCCRPSCCCTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCC 77
QY 145 CRPSCCISCCCKPSCCCTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCC 174
Db 78 CRP-----LCCQTTCRATCCRPSCCGSSC 102

RESULT 9
US-10-184-644-283
; Sequence 283, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 283
; LENGTH: 2294
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-283

Query Match 29.9%; Score 443; DB 9; Length 2294;
Best Local Similarity 31.5%; Pred. No. 2.9e-21;
Matches 96; Conservative 20; Mismatches 85; Indels 104; Gaps 22;
QY 4 SCGGSVCSQSGGAGQESCCRP--SCCCTTCCT-----TCCRSCTISSC 49
Db 333 ACCGAGCC-CCCGCGAGCGACCGTCCACGACCCCTGGCTGGCAGCTTCTCCAGAGCAATC 391
QY 50 CRPSCC--ISSCKPSCCTTCRPSCCISCCRPSCC-----CCR-TTCCRPSCCISCCRPSC 89
Db 392 C---CGGAGACCAACCCCTCTTTGGGCGACTGCTGGACCCCTCTTC--ACC--ACCTTTC 444
QY 90 -----CISCKPSCCTTCRPSCCISCCRPSCC-----ISSCKP 127
Db 445 AGCGAGCGCTCGGCGCCCTCCGCGACCAACCC--TCCGGCGGCGAGACGACCTTCGAGC-- 500
QY 128 SCCTTC-----CRPSCCISCKPSCCISCCRPSCCCTTC-----CRPSCCISCC 174
Db 501 ACTCTCAGCGCGGAGACAGACCGCGCGGACCC-ACCCTTCGACGACCACTGAGAGC 559
QY 175 YRPCCCPSCCRPACCLIS-----SCHPSCCVSSCRCPSC-PTTCRPTTCRPHDCC 225
Db 560 ---GCC-GACCAACCCCTGTAGCGACCAACCCCTACC-GCGCGCCACGACTTC--CC 606
QY 226 GSSCC 230
Db 607 GGACC 611

RESULT 10
US-10-184-634-283
; Sequence 283, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 283
; LENGTH: 2294
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-283


```
: LENGTH: 2236
: TYPE: DNA
: ORGANISM: Homo Sapien
US-10-123-155-381

Query Match      29.4%; Score 437; DB 9; Length 2236;
Best Local Similarity 32.2%; Pred. No. 6e-21;
Matches 89; Conservative 9; Mismatches 110; Indels 68; Gaps 16;

QY 5 CCGSVCDQSCGGLGQESCCRPSCCOTTCCTTCRPSCCISCCRPSCCISCCCKPSC 64
Db ||| |: ||| || || || || || || || || || || || || || ||
QY 40 CCGGCTGGCGCCCGCGCC--GCCGCGCGCCACGCCCAACC---CGGCGCGCGC 94
Db ||| || || || || || || || || || || || || || || || || ||
QY 65 CLTTCCTTCRPSCCISCCRPSC-----ISSCKPSC-----CRTTCRCP 107
Db || || || || || || || || || || || || || || || || || ||
QY 95 C--CCCTAGCC---CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 149
Db || || || || || || || || || || || || || || || || || ||
QY 108 SCCISSCCRPSCCISCCCKPSCRTTCRPSCCISCCRPSCCISCC-----KPSCC 160
Db | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 150 CCGCGCGGAGGCGCGCC---CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGGAACC 206
Db |||| || || || || || || || || || || || || || || || ||
QY 161 -----QTTCCRPSCCISCCRPSCQCPSCCPA--CC-----ISSCHPSCCVS 202
Db |||| || || || || || || || || || || || || || || || ||
QY 207 GGGCGGATCTCGCGGTC-AAACC--ACCTGATCCCATAAACATTCATCC--TCCCG 261
Db || || || || || || || || || || || || || || || || || ||
QY 203 SC-RCPFSCPTT-----CCRTTCFHPICCGSSCC 230
Db | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 GCGGCGCGGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 297
Db || || || || || || || || || || || || || || || || || ||
```

Search completed: June 11, 2003, 11:14:44
Job time : 46 secs


```

QY 101 RTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 160
Db 94 QCCYKRP-CCCSGGCGSCCQSCCKPCCSCCKP-CCSCGGCGSCCQSCCKPCCS 151
QY 161 QTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 200
Db 152 QSSCKPCCSCGCG-SSCCQSCCKPCCSCCKPCCSCCKPCCSCCKPCCSCCKPCC 190

RESULT 2
KR2D_SHEEP
ID KR2D_SHEEP STANDARD: PRT: 181 AA.
AC P08131.
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Keratin, high-sulfur matrix protein, B2D.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8329218; PubMed=6193483;
RA Powell B.C., Sleight M.J., Ward K.A., Rogers G.E.;
RT "Mammalian keratin gene families: organization of genes coding for
RT the B2 high-sulfur proteins of sheep wool.";
RL Nucleic Acids Res. 11:5327-5346(1983).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 KDA, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kda).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X01610; CAA35759.1; -.
CC PIR: S07911; KRSHHD.
CC InterPro: IPR002494; Keratin_B2.
CC Pfam: PF01500; Keratin_B2: 1.
CC Keratin; Repeat.
CC INIT_MET 0
CC DOMAIN 26 85 6 x 10 AA TANDEM REPEATS.
CC REPEAT 26 35 1.
CC REPEAT 36 45 2.
CC REPEAT 46 55 3.
CC REPEAT 56 65 4.
CC REPEAT 66 75 5.
CC REPEAT 76 85 6.
CC SEQUENCE 181 AA; 18679 MW; D3E4874E21757B12 CRC64;

Query Match 29.2%; Score 434; DB 1; Length 181;
Best Local Similarity 37.1%; Pred. No. 5.8e-29;
Matches 76; Conservative 27; Mismatches 54; Indels 48; Gaps 7;

QY 4 SCCG-SVCSQSGGLGQESCCQTCCTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 62
Db 1 ACCSTSGCGPPTSTG---GTGSGNFCQPTCCQTSCTCCOPTSIQTSCTCCQ 56
QY 63 SCCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 104
Db 57 -----TSIQTSCTCCQPTSIQTSCTCCQPTCLQTSCTGTCGIGGSIYGVGVSSGAVSSRTKW 111
QY 105 CRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 159

```

```

Db 112 CRPCDREVECTSLPPCCVWSCTSPSCQLYAASCCRPSCGSCGRPAC-----C 162
QY 160 COTTCRPPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 184
Db 163 CQPTCTEPVC-----EPTCCPTC 181

RESULT 3
KRUA_HUMAN
ID KRUA_HUMAN STANDARD: PRT: 169 AA.
AC P26371.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, ultra high-sulfur matrix protein A (UHS keratin A) (UHS
DE KerA).
GN KRNI OR UHSK1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91115951; PubMed=1703541;
RA McKinnon P.J., Powell B.C., Rogers G.E.;
RT "Structure and expression of genes for a class of cysteine-rich
RT proteins of the cuticle layers of differentiating wool and hair
RT follicles.";
RL J. Cell Biol. 111:2587-2600(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148005; PubMed=10023043;
RA Perez C., Aurio J., Gerst C., Hornard B.A., Egly J.-M.;
RT "Genomic organization and promoter characterization of two human UHS
RT keratin genes.";
RL Gene 227:137-148(1999).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 KDA, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kda).
CC -----
CC -!- TISSUE SPECIFICITY: CUTICLE LAYERS OF DIFFERENTIATING
CC HAIR FOLLICLES.
CC -!- DEVELOPMENTAL STAGE: AT A LATE STAGE OF FIBER DIFFERENTIATION.
CC -!- DOMAIN: MAINLY COMPOSED OF CYS-RICH (CR), GLY-RICH (GR) AND SER-
CC RICH (SR) REPEATS.
CC -!- SIMILARITY: BELONGS TO THE UHS KERATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X55293; CAA39005.1; -.
CC HSSP: P01064; IP12.
CC Genew: HGNC:6409; KRNI.
CC MIM: 148021; -.
CC Keratin; Repeat; Multigene family.
CC SEQUENCE 169 AA; 16276 MW; 219B14FEB49D4AB CRC64;

Query Match 28.4%; Score 420.5; DB 1; Length 169;
Best Local Similarity 46.4%; Pred. No. 6.8e-28;
Matches 81; Conservative 13; Mismatches 56; Indels 25; Gaps 8;

QY 3 SSCGSGVCS--DQSGVGLGQESCC-----CRPSCCQTCCTTCRPPSCCISCCRPSCC 54

```



```

AC P02442;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Keratin, high-sulfur matrix protein, IIA3 (M2.6 protein).
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RC STRAIN=South African angora;
RA Joubert F.J.;
RT "Studies on the high-sulphur proteins of reduced mohair. The
RT isolation and the amino acid sequence of protein SCMK-B2.6.";
RL J. S. Afr. Chem. Inst. 28:250-263(1975).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
DR PIR: A92978; KRGT3J.
DR PIR: B92978; KRGT3M.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
FT VARIANT 23 23 Y -> C (IN MINOR COMPONENT).
FT VARIANT 119 119 MISSING (IN MINOR COMPONENT).
FT SEQUENCE 132 AA; 14255 MW; 6C0CC8FA13AB28B CRC64;
SQ
Query Match 19.1%; Score 284; DB 1; Length 132;
Best Local Similarity 40.1%; Pred. No. 5.5e-17;
Matches 57; Conservative 11; Mismatches 56; Indels 18; Gaps 5;
QY 4 SCGGSVCSDSCGGGLGQESCCR-PSCQTTTCRTTCRPPSCCISCCRPSCCISCCCKP 62
DB 3 SCGGPTFSSLSGGGGLQPRYYRDPCCCRPVSCQTTVSRPVTFVSRCTRP-----ICP 56
QY 63 SCLATTCCTTCRPPSCCISCCRP-SCCISSC-----CKPSCRTTCRPPSCCISCCRP 117
DB 57 -----CRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQPCCRP 110
QY 118 SCCISCCCKPSCRTTCRPPSC 139
DB 111 TSCQAPCSTRTCRTFRTSPCC 132
RESULT 9
KR3A_SHEEP STANDARD; PRT: 130 AA.
AC P02443;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Keratin, high-sulfur matrix protein, IIA3A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RX MEDLINE=74022242; PubMed=4584026;
RA Swart L.S., Haylett T.;
RT "Studies on the high-sulphur proteins of reduced Merino wool. Amino
RT acid sequence of protein SCMK-B3A3.";
RL Biochem. J. 133:641-654(1973).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
CC -!- MISCELLANEOUS: THE SOURCE OF THIS PROTEIN IS MERINO WOOL..
DR PIR: A02840; KRSHA3.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
SQ SEQUENCE 131 AA; 14163 MW; 8C0B56C0070B1C50 CRC64;
Query Match 18.2%; Score 269.5; DB 1; Length 131;
Best Local Similarity 38.3%; Pred. No. 7.9e-16;
Matches 54; Conservative 12; Mismatches 58; Indels 17; Gaps 6;
QY 4 SCGGSVCSDSCGGGLGQESCCR-PSCQTTTCRPPSCCISCCRPSCCISCCCKP 63
DB 3 SCGGPTFSSLSGGGGLQPRYYRDPCCCRPVSCQTTVSRPVTFVSRCTRP 52
QY 64 SCLATTCCTTCRPPSCCISCCRP-SCCISSC-----CKPSCRTTCRPPSCCISCCCKP 118
DB 53 C---EPCRRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQPCCRP 110
QY 119 CCISCCCKPSCRTTCRPPSC 139

```

```

CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
CC -!- MISCELLANEOUS: THE SOURCE OF THIS KERATIN IS MERINO WOOL..
DR PIR: A02841; KRSHA3A.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
SQ SEQUENCE 130 AA; 13894 MW; 8ACF3719FBBB8361 CRC64;
Query Match 18.8%; Score 279; DB 1; Length 130;
Best Local Similarity 39.5%; Pred. No. 1.4e-16;
Matches 58; Conservative 11; Mismatches 54; Indels 24; Gaps 7;
QY 4 SCGGSVCSDSCGGGLGQESCCR-PSCQTTTCRPPSCCISCCRPSCCISCCCKP 63
DB 3 SCGGPTFSSLSGGGGLQPRYYRDPCCCRPVSCQTTVSRPVTFVSRCTRP 53
QY 64 SCLATTCCTTCRPPSCCISCCRP-SCCISSC-----CKPSCRTTCRPPSCCISCCCKP 118
DB 54 C---EPCRRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQPCCRP 111
QY 119 CCISCCCKPSCRTTCRPPSCCISCC 145
DB 112 SC-----PSAPKPTTC--KTFRTSPCC 130
RESULT 10
KR3A_SHEEP STANDARD; PRT: 131 AA.
AC P02441;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Keratin, high-sulfur matrix protein, IIA3A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RX MEDLINE=74022242; PubMed=4584026;
RA Swart L.S., Haylett T.;
RT "Studies on the high-sulphur proteins of reduced Merino wool. Amino
RT acid sequence of protein SCMK-B3A3.";
RL Biochem. J. 133:641-654(1973).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
CC -!- MISCELLANEOUS: THE SOURCE OF THIS PROTEIN IS MERINO WOOL..
DR PIR: A02840; KRSHA3.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
SQ SEQUENCE 131 AA; 14163 MW; 8C0B56C0070B1C50 CRC64;
Query Match 18.2%; Score 269.5; DB 1; Length 131;
Best Local Similarity 38.3%; Pred. No. 7.9e-16;
Matches 54; Conservative 12; Mismatches 58; Indels 17; Gaps 6;
QY 4 SCGGSVCSDSCGGGLGQESCCR-PSCQTTTCRPPSCCISCCRPSCCISCCCKP 63
DB 3 SCGGPTFSSLSGGGGLQPRYYRDPCCCRPVSCQTTVSRPVTFVSRCTRP 52
QY 64 SCLATTCCTTCRPPSCCISCCRP-SCCISSC-----CKPSCRTTCRPPSCCISCCCKP 118
DB 53 C---EPCRRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQPCCRP 110
QY 119 CCISCCCKPSCRTTCRPPSC 139

```

```
Db 111 SCOPAPCSRTTCRTFTSPCC 131
RESULT 11
YAV2_XANCV
ID YAV2_XANCV STANDARD; PRT; 784 AA.
AC P14728;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 82 kDa avirulence protein in AVRBS3 region.
OS Xanthomonas campestris (pv. vesicatoria).
OG Plasmid pXv11.
OC Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 71-21;
RX MEDLINE=89384426; PubMed=2550761;
RA Bonas U., Stall R.E., Staskiewicz B.;
RT "Genetic and structural characterization of the avirulence gene
RL avrBs3 from Xanthomonas campestris pv. vesicatoria.";
RC Mol. Gen. Genet. 218:127-136(1989)
CC -!- DISEASE: THE TERM AVIRULENT DESCRIBES A POTENTIALLY VIRULENT
CC PATHOGEN WHICH IS UNABLE TO INDUCE DISEASE SYMPTOMS IN SPECIFIC
CC CULTIVARS OF PEPPER.
CC -!- MISCELLANEOUS: THIS IS ONE OF THE HYPOTHETICAL PROTEINS CODED BY
CC THE OPEN READING FRAMES WITHIN THE REGION REQUIRED FOR AVRBS3
CC ACTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X16130; ; NOT_ANNOTATED_CDS.
CC DR PIR; J00317;
CC DR InterPro; IPR005042; Avirulence.
CC DR Pfam; PF03377; Avirulence; 1.
CC SW Hypothetical protein; Plasmid; Virulence.
CC KQ SEQUENCE 784 AA; 82074 MW; 3B2624B411793744 CRC64;
Query Match 15.1%; Score 224.5; DB 1; Length 784;
Best Local Similarity 22.6%; Pred. No. 9.1e-12;
Matches 109; Conservative 19; Mismatches 88; Indels 267; Gaps 29;
QY 1 MVSSCCG-----SVCSQSCGQGLGQESCCRP--CCQTTCC----- 34
DB 121 MATTCGVRPWAHSTGNSRCTVSS-----ACLPPLSLAMATTCGVRPWAHSTGN 172
QY 35 -----CRTCC--RP-----SCCISCCRP----- 52
DB 173 SRTVTSAGLPPPLLAMATTCGVRPWAHSTGNSRCTVSSACLPPLSLAMATTCGVRP 232
QY 53 -----SCCISSCCKPSCCLTTCRTTC--CRP-----SCCISCCRP----- 87
DB 233 WAWHSTGNSRCTVSSACLPPLSLAMA--TTTCGVRPWAHSTGNSRCTVSSACLPPLSLA 290
QY 88 -----SCCISSCCKPS--CCRTTC--CRP-----SCCISCC 114
DB 291 MATTCGVRPWAHSTGNSRCTVSSACLPPLLAMATTCGVRPWAHSTGNSACTVSSA 350
QY 115 CRP-----SCCISSCCKPS--CCRTTC--RP----- 137
DB 351 CLPPLLAMATTCGVRPWAHSTGNSRCTVSSACLPPLLAMATTCGVRPWAHSTGN 410
QY 138 -SCCISSCCRP-----SCCISSCCKPS--CCQTTCC--CRP 167
DB 411 SRTVTSAGLPPPLSLAMATTCGVRPWAHSTGNSRCTVSSACLPPLSLAMATTCGVRP 470
```

```
QY 168 -----SCCISSCYRPOCCOP-----SCCRP-----ACCISSCCHP--- 197
DB 471 WAWHSTGNSACTVSSA-----CLPPLLAMATTCGVRPWAHSTGNSACTVSSACLPPL 526
QY 198 -----SCCVSSCRCP---FSCPTTC--CRITCFHP1-----CCGS 227
DB 527 LAMATTCGVRPWAHSTGNSACTVSSACLPPLLAMATTCGVRPWAHSTGNSRCTVSS 586
QY 228 SCC 230
DB 587 SAC 589
RESULT 12
MCS_MOUSE STANDARD; PRT; 197 AA.
AC P15265;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sperm mitochondrial capsule selenoprotein (MCS).
GN MCSP OR MCS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93039675; PubMed=1418626;
RA Karimpour I., Cutler M., Shih D., Smith J., Kleene K.C.;
RT "Sequence of the gene encoding the mitochondrial capsule
RT selenoprotein of mouse sperm: identification of three in-phase TGA
RT selenocysteine codons.";
RL DNA Cell Biol. 11:693-699(1992).
RN [2]
RP SEQUENCE OF 55-197 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=90152148; PubMed=2303168;
RA Kleene K.C., Smith J., Bozorgzadeh A., Harris M., Hahn L.,
RA Karimpour I., Gerstel J.;
RT "Sequence and developmental expression of the mRNA encoding the
RT seleno-protein of the sperm mitochondrial capsule in the mouse.";
RL Dev. Biol. 137:395-402(1990).
CC -!- FUNCTION: STRUCTURAL PROTEIN OF THE SPERM MITOCHONDRIAL CAPSULE.
CC IMPORTANT FOR THE MAINTENANCE AND STABILIZATION OF THE CRESCENT
CC STRUCTURE OF THE SPERM MITOCHONDRIA.
CC -!- SUBCELLULAR LOCATION: KERATINOUS MITOCHONDRIAL CAPSULE.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88463; AAB08438.1; -.
CC DR EMBL; M29603; AAA53045.1; -.
CC DR PIR; A37199; A37199.
CC DR MGD; MGI:96945; Mcsp.
CC KW Mitochondrion; Selenium; Selenocysteine; Sperm; Testis;
CC Spermatogenesis.
CC FT SE_CYS 7
CC FT SE_CYS 17
CC FT SE_CYS 34
CC FT SE_CYS 34
CC SQ SEQUENCE 197 AA; 21015 MW; 721R84F7129F02B1 CRC64;
Query Match 14.9%; Score 221.5; DB 1; Length 197;
Best Local Similarity 33.9%; Pred. No. 7.2e-12;
```

Matches 61; Conservative 9; Mismatches 63; Indels 47; Gaps 15;

Qy 24 CC-RPS---CC---QTTCCRTTCRPSCCISCCRPSCCISCCRPSCCLTTCRTTCR 76
 Db 6 CCGRPSISCCNTQETTFENTOVKCCCTKSKRKNVCARSPTLKKMSDPSKTNQCP 65

Qy 77 PSSC-SSCCRPSCISCC--KPSCC-RTTCRPSCCISCCRPSCCISCCRPSC-- 129
 Db 66 PPCCPPKPCPP-----KPCQPKRPPCCPKSPCCPKP---SPCCPKPC---PCPPPCPP 115

Qy 130 CRTTC---CRPSCCISCCRPSCCISCCRPSCCQTTTC--RPSCCISCCYRPPCCQPSCC 185
 Db 116 CPATCPCLPKPCPCQKC-----SCCPKKC---TCCPPPPC-----CAQPTCC 156

RESULT 13

PSIA_DICDI
 ID PSTA_DICDI STANDARD; PRT: 1046 AA.
 AC P11976;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Prestalk protein precursor.
 GN ECMB.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RC STRAIN=AX2;
 RX MEDLINE=88015608; PubMed=3658700;
 RA Ceccarelli A., McRobbie S.J., Jermyn K.A., Duffy K., Early A.,
 Williams J.G.;
 RT "Structural and functional characterization of a Dictyostelium gene
 encoding a DIF inducible, prestalk-enriched mRNA sequence";
 RL Nucleic Acids Res. 15:7463-7476(1987).
 RN [2]

RP SEQUENCE OF 1-17 FROM N.A.
 RX MEDLINE=87257883; PubMed=3600646;
 RA Ayres K., Neuman W., Rowekamp W.G., Chung S.;
 RT "Developmental regulation of DNase I-hypersensitive sites in
 Dictyostelium discoideum";
 RL Mol. Cell. Biol. 7:1823-1829(1987).
 CC -!- INDUCTION: BY THE PUTATIVE STALK-SPECIFIC MORPHOGEN DIF
 CC -!- SIMILARITY: THE REPEATED SEQUENCES ARE HIGHLY HOMOLOGOUS TO A
 CC REPEAT OF THE PDD63 MRNA OF DICTYOSTELIUM DISCOIDEUM.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M16345; AAA33216.1;
 DR PIR; A26838; A26838.
 DR HSSP; P15445; 1A3F.
 DR DictyDb; DD03006; ecmb.
 DR InterPro; IPR001673; S_mold_repeat.
 DR Pfam; PF00526; S_mold_repeat; 42.
 DR ProDom; PD006869; S_mold_repeat; 8.
 KW Signal; Repeat.
 FT SIGNAL 1 18 PROBABLE.
 FT CHAIN 19 1046 PRESTALK PROTEIN.
 FT DOMAIN 31 1012 41 X 24 AA TANDEM REPEATS, CYS-RICH.
 FT REPEAT 31 53 X-1.
 FT REPEAT 54 77 X-2.
 FT REPEAT 78 101 X-3.
 FT REPEAT 102 125 A-1.
 FT REPEAT 126 149 A-2.
 FT REPEAT 150 173 B-1.

FT REPEAT 174 197 A-3.
 FT REPEAT 198 221 A-4.
 FT REPEAT 222 245 B-2.
 FT REPEAT 246 269 A-5.
 FT REPEAT 270 293 A-6.
 FT REPEAT 294 317 H-3.
 FT REPEAT 318 341 A-7.
 FT REPEAT 342 365 A-8.
 FT REPEAT 366 389 H-4.
 FT REPEAT 390 413 A-9.
 FT REPEAT 414 437 A-10.
 FT REPEAT 438 461 H-5.
 FT REPEAT 462 485 A-11.
 FT REPEAT 486 509 A-12.
 FT REPEAT 510 533 B-6.
 FT REPEAT 534 557 A-13.
 FT REPEAT 558 581 X-4.
 FT REPEAT 582 605 A-14.
 FT REPEAT 606 629 A-15.
 FT REPEAT 630 653 A-16.
 FT REPEAT 654 677 H-7.
 FT REPEAT 678 701 A-17.
 FT REPEAT 702 725 A-18.
 FT REPEAT 726 749 H-8.
 FT REPEAT 750 773 A-19.
 FT REPEAT 774 797 A-20.
 FT REPEAT 798 821 H-9.
 FT REPEAT 822 845 A-21.
 FT REPEAT 846 869 A-22.
 FT REPEAT 870 893 A-23.
 FT REPEAT 894 917 H-10.
 FT REPEAT 918 941 A-24.
 FT REPEAT 942 964 A-25.
 FT REPEAT 965 988 A-26.
 FT REPEAT 989 1012 A-27.
 SQ SEQUENCE 1046 AA; 108823 MW; F9D6C04FD7D85818 CRC64;

Query Match 14.7%; Score 217.5; DB 1; Length 1046;
 Best Local Similarity 21.7%; Pred. No. 3.9e-11;
 Matches 78; Conservative 41; Mismatches 106; Indels 135; Gaps 24;

Qy 3 SSSC-----GSVCSQSCGGGIG-----QESCCRPSCOTT-CGRT--TC-C 40
 Db 641 TGCCHTPIVDDHNACTEDKCTQSGVHTPIACDDKNACTVDSGNSGCGCHTPIASD 700

Qy 41 RPSCCISCCRPSCCI-----SSCC-----KPSCCITTCGRT 72
 Db 701 NNACTVDSGNSGCGVHTPIVDDHNACTEDKCTONGGVHTPIACDDKNACTVDSGNS 760

Qy 73 T-CGRP-----SCCISCCRPSCCI-----SSCCRPSCRT-----TCGR--- 106
 Db 761 TGCCHTPIKCDNNHNPCTVDSGNSGCGVHTPIVDDHNACTEDKCTQNGVHTPIACDD 820

Qy 107 -PSCCISCCRPSCCI-----SSCCRPSCRTT-CGRP--SCCISCCRPSCIS 153
 Db 821 LNSCTADSCNSGCGVHTPIVDDHNACTEDKCTQNGVHTPIVDDHNPCTVDSGNS 880

Qy 154 -CC-----KPSCCOTTTCRP-----SCCISCCYRPPCC--OBSCC-- 185
 Db 881 TGCCHTPIVDDHNACTEDKCTQSGVHTPIVDDHNACTVDSGNSGCGCHTPIACDD 940

Qy 186 RPACCCISCCRPSCCGVSCRPSCP-----TTCCRTTCFH-PI-----CCGSSCC 230
 Db 941 NNPCTIDSCNSGCGVHT---PMNCDGNGNFCTLDSCCGTGTCTHTPIIIDNNPCTVDSG 997

RESULT 14
 ID STC_DROME
 AC P40798;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Shuttle craft protein.
GN STC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=96104568; PubMed=8524296;
RA Strombakis N.D., Li Z., Tolia P.P.;
RT "A homolog of human transcription factor NF-X1 encoded by the
RT Drosophila shuttle craft gene is required in the embryonic central
RT nervous system.";
RL Mol. Cell. Biol. 16:192-201(1996).
CC -|- FUNCTION: PLAYS AN ESSENTIAL ROLE DURING THE LATE STAGES OF
CC EMBRYONIC NEUROGENESIS. MAY EITHER FINE-TUNE THE GUIDANCE OF THE
CC SPATIAL MAINTENANCE OF THE MIGRATING SNB AND IN NERVE ROOTS, OF
CC WHICH ARE COMPOSED OF AXONS ORIGINATING FROM DISTINCT GROUPS OF
CC MOTOR NEURONS AND MAY BE REQUIRED TO EITHER GUIDE OR MAINTAIN THE
CC POSITION OF THESE NERVES ALONG A DIRECT AND STRAIGHT PATH TO THEIR
CC ULTIMATE TARGETS IN PARTICULAR MUSCLE FIELDS. MAY PLAY A ROLE IN
CC EGG CHAMBER DEVELOPMENT AND/OR MAY CONFER ESSENTIAL MATERNAL
CC CONTRIBUTIONS TO THE EARLY EMBRYO.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- TISSUE SPECIFICITY: OVARIES AND EMBRYONIC CENTRAL NERVOUS SYSTEM.
CC -|- DEVELOPMENTAL STAGE: MAJOR EXPRESSION IS SEEN IN THE OVARIES WHILE
CC MODERATE LEVELS OF EXPRESSION ARE OBSERVED DURING EMBRYOGENESIS
CC AND THROUGHOUT SUBSEQUENT STAGES OF FLY DEVELOPMENT.
CC -|- SIMILARITY: TO YEAST YNL023C AND HUMAN NF-X1.
CC -|- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09306; AAB60255.1; -;
DR TRANSFAC; T01688; -;
DR FlyBase; FBgn0001978; stc.
DR InterPro; IPR001374; R3H.
DR InterPro; IPR000967; Znf_NFX1.
DR InterPro; IPR001841; Znf_Ring.
DR Pfam; PF01422; zf-NF-X1; 8.
DR Pfam; PF01424; R3H; 1.
DR SMART; SM00393; R3H; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00438; Znf_NFX; 9.
DR PROSITE; PS01359; ZF_PHD.1; 1.
DR PROSITE; PS00518; ZF_RING.1; 1.
DR PROSITE; PS50089; ZF_RING.2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;
KW Alternative splicing; RNA-binding; Zinc-finger.
FT DOMAIN 241 268
FT ZN_FING 386 433
FT ZN_FING 464 500
FT REPEAT 464 500
FT REPEAT 517 554
FT REPEAT 575 614
FT REPEAT 634 675
FT REPEAT 723 762
FT REPEAT 834 871
FT REPEAT 875 904
FT VARSPLIC 109 115
SQ SEQUENCE 1106 AA; 123261 MW; 8CE146DB37BB8D65 CRC64;
Query Match 14.2%; Score 210; DB 1; Length 1106;
Best Local Similarity 22.4%; Pred. No. 1.6e-10;
Matches 76; Conservative 22; Mismatches 120; Indels 122; Gaps 15;

QY 6 CGSVCSQD-QSCGGGLGQESCCRPSC-----OTTCR 36
DB 517 CGEICDKLLNGEHRQ-QAFCHSGKCAACSFVVQOCHGCKQERKVPCTRESQKRTYSCK 576
QY 37 TTCRPSCCJSSCCRPSCCLSSC--CKPSCCLTTC-----RTTCRRV-SGCLIS 82
DB 577 DSCGQPLPCGHHKCKDSCHAGSCRPCKLSPQITSCPCGKIPVPAGORSCLDIPITCKG 636
QY 83 SSCR-----PSCCLSSC-----CKPSC-----C 100
DB 637 ICSRTLRCGKPAHPHQVGSCKHLGOCPCPKOTGVKRCGCHMDOMIKKCOLNRAADARC 696
QY 101 RTTCRPSCCJSSCCRPSC--ISSCKPSCCCTTCRPSCCJSSCCRPSC--CISCK 156
DB 697 KRCTKRKSGKHKVNECCIDIDHHCPLPCNRTLSGKHKCDQPCIRGNCPICYKSSFE 756
QY 157 PSCCOTTCRPSCCJSSCYRQCC--ONSCCRPACCLSSC-----CH-----PSCKV 201
DB 757 ELYCE-----CGAEVYPPVPCGKPKICKLPSSRIHPCDHPPOHCHSGTCTPCMI 809
QY 202 SSCR-----CPFSCTTCCTTCRTTCRPHICCGSSCC 230
DB 810 FTKLCHGNHWRKTIPTCSUPNFCGMAJCKPLPCGGHKC 849
RESULT 15
NFX1_HUMAN STANDARD; PRT; 1104 AA.
AC Q12986;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional repressor NF-X1 (Nuclear Transcription factor, X box-
DE binding, 1).
GN NFX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95053707; PubMed=7964459;
RA Song Z., Krishna S., Thianos D., Strominger J.L., Ono S.J.;
RT "A novel cysteine-rich sequence-specific DNA-binding protein
RT interacts with the conserved X-box motif of the human major
RT histocompatibility complex class II genes via a repeated Cys-His
RT domain and functions as a transcriptional repressor.";
RL J. Exp. Med. 180:1763-1774(1994).
CC -|- FUNCTION: REPRESSOR OF HLA-DRA TRANSCRIPTION. BINDS TO THE X-BOX
CC MOTIF OF CLASS II MHC GENES. MAY PLAY AN IMPORTANT ROLE IN
CC REGULATING THE DURATION OF AN INFLAMMATORY RESPONSE BY LIMITING
CC THE PERIOD IN WHICH CLASS II MHC MOLECULES ARE INDUCED BY
CC INTERFERON-GAMMA.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- INDUCTION: BY INTERFERON GAMMA.
CC -|- SIMILARITY: TO D.MELANOGASTER SHUTTLE CRAFT PROTEIN (STC) AND
CC YEAST YNL023C.
CC -|- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15306; AAA69517.1; -;
DR TRANSFAC; T01694; -;
DR Genew; HGNC:7803; NFX1.
DR MIM; 603255; -;
DR InterPro; IPR001374; R3H.

```
DR InterPro: IPR000967; Znf_NFX1.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_Ring.
DR Pfam: PF01422; Zf-NF-X1; 8.
DR Pfam: PF01424; R3H; 1.
DR SMART: SM00393; R3H; 1.
DR SMART: SM00184; RING; 1.
DR SMART: SM00438; Znf_NFX; 9.
DR PROSITE: PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE: PS00016; ZF_PHD_2; 1.
DR PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE: PS00089; ZF_RING_2; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Repeat; Zinc-finger.
FT ZN_FING 342 393 RING-TYPE.
FT DOMAIN 424 876 7 X APPROXIMATE REPEATS, CYS-RICH.
FT REPEAT 424 463 1.
FT REPEAT 480 517 2.
FT REPEAT 541 580 3.
FT REPEAT 606 647 4.
FT REPEAT 695 734 5.
FT REPEAT 806 842 6.
FT REPEAT 843 876 7.
SQ SEQUENCE 1104 AA; 123149 MW; 36FE4292F78130C CRC64;

Query Match 13.3%; Score 198; DB 1; Length 1104;
Best Local Similarity 21.9%; Pred. No. 1.5e-09;
Matches 80; Conservative 18; Mismatches 118; Indels 150; Gaps 19;

Qy 6 CGSVCSQSCG-----QGLQESC-----CRPSCCQ-----TT 33
Db 517 CGSTRDVLGTDVGKSGDFSLCTGKDLKCGNHTCSQVCHPOPCOCPRLQLVR 576
Qy 34 CC-----RTCCRP-SCCISCCRPSCC-----ISSC-----CKP 62
Db 577 CCPCGQTPLSOLLELGLSSSRKTCMDPVPSGKVGCKPLPCGSLDFIHTCEKLCHEGDCGP 636
Qy 63 -----SC-----C-----LTTCRTTCCRPSCCISCCRPSCCIS--CCK 96
Db 637 VSRTSVISCRCGSRFTKELPCTSLKSEDATFMDKRCNKNKRLCGRHKCNEICCVDEHKCP 696
Qy 97 PSQCRITCCRPSCCISCCRPSCCISCCCKPSCCCTTC-----CRPSCCIS 142
Db 697 LNCGRKLCGLHRCCEPCHRGNC--QTCWQASFDLTCGASVIYPPVPCGTRPPECTQ 754
Qy 143 SCCR-----PSC-----CIS-----SCCRPSCCOT 162
Db 755 TCARVHCDHPVYHSGHSEKCPCTFLTQKWCWKHEFRSNIPCHLVDISGLP-CSAT 813
Qy 163 TCCRPSCCISCYRPCCQCPSCCRPACCI--SSCHPSCCVSSCRCPFCPTTCRTTCF 220
Db 814 LPCGMHKKQRLCHKGCELVDEPCQPCCTTPRADCGHP--CNAPCHTSSPCPVACKAKVE 871
Qy 221 HPICCG 226
Db 872 LQCECG 877
```

Search completed: June 11, 2003, 11:04:58
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 11:02:42 ; Search time 72 Seconds
(without alignments)
425.662 Million cell updates/sec

Title: US-09-874-062-3

Perfect score: 1484

Sequence: 1 MVSSCGSVCSQSGGGLG.....PTTCRTTCFHPICGSSCC 230

Scoring Table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1484	100.0	230	22	AAE02058 Human four disulfide
2	1218.5	82.1	206	22	ABG29368 Novel human diago
3	767	51.7	301	22	AAW79986 Human protein SEQ
4	758.5	51.1	250	22	AAW79002 Human protein SEQ
5	599	40.4	120	22	ABG12277 Human hair keratin
6	552.5	37.2	256	22	ABG08337 Novel human diago
7	552.5	37.2	351	22	ABG23698 Novel human diago
8	523.5	35.3	267	22	ABG27981 Novel human diago
9	513	34.6	231	22	ABG32110 Peptide #4761 enco
10	513	34.6	231	22	ABG37362 Peptide #4868 enco

11	513	34.6	231	22	ABG22648 Protein #4647 enco
12	513	34.6	231	22	AAW58051 Human brain expres
13	513	34.6	231	22	AAW70492 Human bone marrow
14	513	34.6	231	22	AAW18328 Peptide #4762 enco
15	513	34.6	231	22	AAW30815 Peptide #4852 enco
16	513	34.6	231	22	AAW05935 Peptide #4617 enco
17	513	34.6	231	22	ABG40131 Human peptide enco
18	466.5	31.4	134	22	ABG39521 Peptide #7027 enco
19	466.5	31.4	134	22	ABG24266 Peptide #6265 enco
20	466.5	31.4	134	22	AAW60215 Human brain expres
21	466.5	31.4	134	22	AAW72838 Human bone marrow
22	466.5	31.4	134	22	AAW19759 Peptide #6193 enco
23	466.5	31.4	134	22	AAW33068 Peptide #7105 enco
24	466.5	31.4	134	22	ABG42667 Human peptide enco
25	448	30.2	102	22	ABG38806 Peptide #6312 enco
26	448	30.2	102	22	AAW59445 Human bone marrow
27	448	30.2	102	22	AAW72006 Human bone marrow
28	448	30.2	102	22	AAW32278 Peptide #6315 enco
29	448	30.2	102	22	ABG41821 Human peptide enco
30	438.5	29.5	229	22	AAW79404 Human protein SEQ
31	408.5	27.5	171	23	AAE18324 Sheep keratin-asso
32	405	27.3	215	22	ABG26428 Novel human diago
33	394.5	26.6	1329	23	AAU91279 Human Nw3a protei
34	393	26.5	398	22	ABG19008 Novel human diago
35	372	25.1	222	22	AAW41252 Human polypeptide
36	369.5	24.9	167	22	AAW39466 Human polypeptide
37	362.5	24.4	1679	22	AAU07343 1-aminocyclopropan
38	348.5	23.5	85	22	ABG40077 Peptide #7583 enco
39	348.5	23.5	85	22	ABG24566 Protein #6565 enco
40	348.5	23.5	85	22	AAW60833 Human brain expres
41	348.5	23.5	85	22	AAW73514 Human bone marrow
42	348.5	23.5	85	22	AAW19995 Peptide #6429 enco
43	348.5	23.5	85	22	AAW33710 Peptide #7747 enco
44	348.5	23.5	85	23	ABG43374 Human peptide enco
45	346.5	23.3	128	21	AAW44987 Human epidermal pr

ALIGNMENTS

RESULT 1
AAE02058
ID AAE02058 standard; Protein: 230 AA.
XX
AC AAE02058;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human four disulfide core domain (FDCD)-containing protein.

Human; four disulfide core domain; FDCD; immunosuppressive; cytostatic; antiarthritic; antirheumatic; antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; dermatological; antinflammatory; vulnerary; antiarteriosclerotic; gene therapy; skin disorder; congenital disorder; mole; freckle; haemangioma; integumentary tumour; basal cell carcinoma; keratosis; melanoma; atherosclerosis; urticaria; photosensitivity; eczema; autoimmune disorder; lupus erythematosus; scleroderma; keloid; stria; erythema; petechia; purpura; xanthelasma; autoimmune disease; rheumatoid arthritis; hyperproliferative disorder; cardiac arrest; cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia; neoplasm; angiodysplasia; nervous system disorder; Alzheimer's disease; infection; ocular disorder; corneal infection.

XX Homo sapiens.
OS WO200140249-A1.
XX
XX
XX
PD 07-JUN-2001.
XX
XX
PF 29-NOV-2000; 2000WO-US32462.
XX
XX 01-DEC-1999; 99US-0168229.

XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Shi Y;
PI WPI: 2001-356153/37.
XX N-PSDB: AAD06303.
DR
XX
XX
PT Four disulfide core domains (FDCD) containing polypeptide and its
PT polynucleotide are used to prevent, treat or ameliorate a medical
PT condition associated with FDCD e.g. skin disorders -
XX
XX Claim 11: Page 230; 235pp; English.
XX
XX The present sequence is four disulfide core domain (FDCD)-
CC containing protein from human. FDCD DNA and protein are used
CC to prevent, treat or ameliorate skin disorders such as congenital
CC disorders e.g. moles, freckles and haemangiomas as well as
CC integumentary tumours (keratosis, basal cell carcinoma, malignant
CC melanoma), injuries and inflammation of the skin, atherosclerosis,
CC urticaria, eczema, photosensitivity, autoimmune disorders (lupus
CC erythematosus, scleroderma), keloids, striae, erythema, petechiae,
CC purpura and xanthelasma. They are also used in the diagnosis and
CC treatment of diseases and disorders of hair and epithelium,
CC autoimmune diseases (e.g. rheumatoid arthritis), hyperproliferative
CC disorders (e.g. neoplasms of the breast or liver), cardiovascular
CC disorders (e.g. cardiac arrest), cerebrovascular disorders (e.g.
CC cerebral ischaemia, angiodysplasia), nervous system disorders (e.g.
CC Alzheimer's disease), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infection). The FDCD-
CC containing proteins are used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. They are used in
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition. The antibodies to FDCD protein are used in alleviating
CC symptoms associated with the disorders and in diagnostic immunoassays
CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC The FDCD sequences are also used in gene therapy.
XX
XX Sequence 230 AA:
SQ
Query Match 100.0%; Score 1484; DB 22; Length 230;
Best Local Similarity 100.0%; Pred. No. 2.9e-94;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVSSCGSVCSQDQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISSCC 60
Db 1 MVSSCGSVCSQDQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISSCC 60
Qy 61 KPSCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 120
Db 61 KPSCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 120
Qy 121 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 180
Db 121 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 180
Qy 181 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 230
Db 181 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 230
RESULT 2
ABG29368
ID ABG29368 standard; Protein: 206 AA.
XX
XX AC ABG29368;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #29359.
XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR
XX N-PSDB: AAS93555.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID NO 59727; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 206 AA:
Query Match 82.1%; Score 1218.5; DB 22; Length 206;
Best Local Similarity 84.3%; Pred. No. 3.2e-76;
Matches 194; Conservative 0; Mismatches 1; Indels 35; Gaps 1;
Qy 1 MVSSCGSVCSQDQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISSCC 60
Db 12 MVSSCGSVCSQDQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISSCC 42
Qy 61 KPSCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 120
Db 43 -----TTCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 96
Qy 121 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 180
Db 97 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 156
Qy 181 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 230
Db 157 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 206

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS72524.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID No 38696; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 256 AA;

Query Match 37.2%; Score 552.5; DB 22; Length 256;
Best Local Similarity 40.6%; Pred. No. 7.7e-31;
Matches 101; Conservative 29; Mismatches 60; Indels 59; Gaps 11;
QY 9 VCSQSCGGGLGQESC-----CRPSCGTTTCRTTCRPPSCGTCRPPSCGTCISS-- 58
DB 39 VCLPGSC-----DSCSDSMQVDDCPESCCEPPCCAPSCCAPAPCLTLVCTPVSCVSSPC 92
QY 59 ---CCRPSCGTTTCRTTCRPPSCGTCISSCRPSCGTCRPPSCGTCISSCC 115
DB 93 CQACEPSPCQSGC--TSSCTPSCCQSSCQACCTSSPCQACCPVCCPKVCCPVCC 150
QY 116 RPSCGTCISSCKP-----SCRTTCRPPSCGTCISSCRPSCGTCISSCRPSCGTCISSCC 168
DB 151 KPVCCKPICCPVCSGASSCCQSSQROPCACCTTSCRPSSSVSLLCRP-VCRSTCCVP- 208
QY 169 CQISSCVRP-QCCQPCRCRPPACGTCISSCRPSCGTCISSCRPSCGTCISSCC 226
DB 209 --IPSCAPASTCQPCRCRPPASVSLLCRPTCSKLSA-----CCG 247
QY 227 -----SSC 229
|||

DB 248 LSSGQKSSC 256

RESULT 7

ABG23698

ID ABG23698 standard; Protein; 351 AA.

XX ABG23698;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #23689.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS87885.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID No 54057; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 351 AA;

Query Match 37.2%; Score 552.5; DB 22; Length 351;
Best Local Similarity 42.4%; Pred. No. 9.6e-31;
Matches 97; Conservative 28; Mismatches 59; Indels 45; Gaps 9;
QY 21 QESCCRPSCGTTTCRTTCRPPSCGTCRPPSCGTCISS-----CCRPSCGTTTCRTTC 75
DB 148 QVDACPESCGEPCCGATSCCAPAPCLTLVCTPVSCVSSPCQACEPSCQSGC--TSSC 205

OY 76 RPSCCISCCRPSCCISCKKPCSCRTTCCRPSCCISCCRPSCCISCKCP-----S 128
 DB 206 TPSCCQSSCOPACTSSPCQACCPVCCPKPCVCCPKPCVCCPKPCVCCGASS 265
 OY 129 CCRTTCCRPSCCISCCRPSCCISCKKPCSCRTTCCRPSCCISCKCP-----S 187
 DB 266 CQSSRRQPCACTTSCRRPSSVLLCRP-VCRSTCCVP---IPSCAPASTCQPCCRP 321
 OY 188 ACCISCCCHPSCC-VSSCRCPFCPTTCCRTTCFHPICCG-----SSC 229
 DB 322 ASCVSLICRPTCSRLSSA-----CCGLSSGOKSSC 351

RESULT 8
 ABG27981
 ID ABG27981 standard; Protein: 267 AA.
 AC ABG27981;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #27972.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS92168.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 58340; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 267 AA:
 Query Match 35.3%; Score 523.5; DB 22; Length 267;
 Best Local Similarity 36.4%; Pred. No. 7.5e-29;
 Matches 99; Conservative 32; Mismatches 68; Indels 73; Gaps 13;
 OY 4 SCCGSGSD-----USCGUGLQSCCKPSCC-----OTTCKRTTCKRPS 43
 DB 23 SICSSACTDSWRVDDC-----PESCCPECCAPAPSLTLVCTPVSCVSSPCCU-TATEPS 76
 OY 44 CCIS---SCCRPSCCISCKKPCSCRTTCCRPSCCISCKCP-----RPSGCCISNVC 95
 DB 77 ACQSGYTSSCTPCYQSSCQPCCTSSPCQACCPVCCPVCCPVCCPVCCPVCCPVCC 136
 OY 96 K--PSCCRTTCCRPSCCISCCRPSCCISCKCP-----PSCCRTTCCRPSCC 141
 DB 137 ESSPSCCQSSCQPCCTTCTSSPCQACCPVCCPVCCPVCCPVCCPVCCPVCCPVCC 196
 OY 142 SSCCRPSCCISCKKPCSCRTTCCRPSCCISCKCP-----PSCCRTTCCRPSCC 199
 DB 197 ASCCRPSSVSLICHP-VCKSTCCVPVPCGASA-----SSCQPCCRTASVSLICRPVC 251
 OY 200 CVSSCRCPFCPTTCCRTTCFHPICCG--SSC 229
 DB 252 SRPAC-----YSLCSGOKSSC 267

RESULT 9
 ABB32110
 ID ABB32110 standard; Peptide: 231 AA.
 XX ABB32110;
 AC ABB32110;
 DT 01-FEB-2002 (first entry)
 DE Peptide #4761 encoded by breast cell single exon nucleic acid probe.
 KW Human: microarray; single exon probe; gene expression; breast;
 KW disease; cancer.
 XX Homo sapiens.
 OS
 XX WO200157271-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00662.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI: 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes.
 XX
 PS Claim 27; SEQ ID NO 15078; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and Bt 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label

CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 231 AA;

Query Match 34.6%; Score 513; DB 22; Length 231;
 Best Local Similarity 40.6%; Pred. No. 3.5e-28;
 Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;

QY 29 CCOTTCCTTCCTCCRPSCCISCCRPSCCISCCRPSCCLTTCCRTTCRPSCCISCCRPSC 88
 DB 1 CCQPVGSEATSCQPVLCVPTSCQPVLCCKSCCQ-----VCEPSCCSAVCTLPS 50
 QY 89 CCISCCCKPSCCTTCRPSCCISCCRPSCCISCCCKPSCCTTCRPSCCIS-----S 143
 DB 51 SCQPVVCEPSCCQPVCTPTCSVTSSCOAVCCDPSPCEPSCSSESSICQATCVALVCEPV 110
 QY 144 CCRPSCCISCCRPSCCQTTTCRPSCCISCCRPSCCISCCRPSCCQCCRPSCCISCCHPSC----- 199
 DB 111 CLRPVCCVQSSCEPPSPVSTCQEPSCCVSSICQICSEPPSPSPAVCVSSPCQPTCYVVK 170
 QY 200 -CVSSCRCPSCPTTCCTTCCTTCFPICGSSCC 230
 DB 171 RCPSPVCEPVPSCPSTSCRPLSCSPGSSASATC 202

RESULT 10
 ABB37362
 ID ABB37362 standard; Peptide; 231 AA.
 XX
 AC ABB37362;
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #4868 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human foetal liver -
 XX
 PS Claim 27; SEQ ID NO 29997; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 231 AA;

Query Match 34.6%; Score 513; DB 22; Length 231;
 Best Local Similarity 40.6%; Pred. No. 3.5e-28;
 Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;

QY 29 CCOTTCCTTCCTCCRPSCCISCCRPSCCISCCRPSCCLTTCCRTTCRPSCCISCCRPSC 88
 DB 1 CCQPVGSEATSCQPVLCVPTSCQPVLCCKSCCQ-----VCEPSCCSAVCTLPS 50
 QY 89 CCISCCCKPSCCTTCRPSCCISCCRPSCCISCCCKPSCCTTCRPSCCIS-----S 143
 DB 51 SCQPVVCEPSCCQPVCTPTCSVTSSCOAVCCDPSPCEPSCSSESSICQATCVALVCEPV 110
 QY 144 CCRPSCCISCCRPSCCQTTTCRPSCCISCCRPSCCISCCRPSCCQCCRPSCCISCCHPSC----- 199
 DB 111 CLRPVCCVQSSCEPPSPVSTCQEPSCCVSSICQICSEPPSPSPAVCVSSPCQPTCYVVK 170
 QY 200 -CVSSCRCPSCPTTCCTTCCTTCFPICGSSCC 230
 DB 171 RCPSPVCEPVPSCPSTSCRPLSCSPGSSASATC 202

RESULT 11
 ABB22648
 ID ABB22648 standard; Protein; 231 AA.
 XX
 AC ABB22648;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #4647 encoded by probe for measuring heart cell gene expression.
 XX
 KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488899/53.
 XX

DR WPI; 2001-488900/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow.
 XX
 PS Example 4; SEQ ID NO: 30799; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 SQ Sequence 231 AA;
 Query Match 34.6%; Score 513; DB 22; Length 231;
 Best Local Similarity 40.6%; Pred. No. 3.5e-28;
 Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;
 QY 29 CCQTTCCRTTCCRPSCGCISSCCRPSCGCISSCKPSCCLTTCRTTCCRPSCGCISSCCRPSS 88
 DB 1 CCQPVGSEATSCQPVLCVPTSCQPVLCSSCCQP-----VWCEPSCSAVCTLPS 50
 QY 89 CCISSCKKPSCCRTTCCRPSCGCISSCCRPSCGCISSCKPSCRTTCCRPSCGCISS 143
 DB 51 SCQPVCEPSCCQPVCTPTTCVTSSCAVCCDPSCEPSCSESSICQATCVALVCEPV 110
 QY 144 CCRPSCGCISSCKKSCCCTTCCRPSCGCISSCCRPSCGCISSCKPSCGCISSCKPSC 199
 DB 111 CLRPPVCCVQSCEPSPVSTQCEPSCGCISSCKPSCGCISSCKPSCGCISSCKPSC 170
 QY 200 -CVSSCRCPSCPTTCCRTTCCFPICCGSSCC 230
 DB 171 RCPSPVCEPSPSTSCRLSCSPGSSASAIC 202
 RESULT 14
 AAM18328
 ID AAM18328 standard; Protein; 231 AA.
 XX
 AC AAM18328;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #4762 encoded by probe for measuring cervical gene expression.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 27; SEQ ID NO 23154; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 231 AA;
 Query Match 34.6%; Score 513; DB 22; Length 231;
 Best Local Similarity 40.6%; Pred. No. 3.5e-28;
 Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;
 QY 29 CCQTTCCRTTCCRPSCGCISSCCRPSCGCISSCKPSCCLTTCRTTCCRPSCGCISSCCRPSS 88
 DB 1 CCQPVGSEATSCQPVLCVPTSCQPVLCSSCCQP-----VWCEPSCSAVCTLPS 50
 QY 89 CCISSCKKPSCCRTTCCRPSCGCISSCCRPSCGCISSCKPSCRTTCCRPSCGCISS 143
 DB 51 SCQPVCEPSCCQPVCTPTTCVTSSCAVCCDPSCEPSCSESSICQATCVALVCEPV 110
 QY 144 CCRPSCGCISSCKKSCCCTTCCRPSCGCISSCCRPSCGCISSCKPSCGCISSCKPSC 199
 DB 111 CLRPPVCCVQSCEPSPVSTQCEPSCGCISSCKPSCGCISSCKPSCGCISSCKPSC 170
 QY 200 -CVSSCRCPSCPTTCCRTTCCFPICCGSSCC 230
 DB 171 RCPSPVCEPSPSTSCRLSCSPGSSASAIC 202
 RESULT 15
 AAM30815
 ID AAM30815 standard; Protein; 231 AA.
 XX
 AC AAM30815;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #4852 encoded by probe for measuring placental gene expression.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Kun on: June 11, 2003, 11:02:46 ; Search time 27 Seconds
(without alignments)
250.640 Million cell updates/sec

Title: US-09-874-062-3
 perfect score: 1484
 Sequence: 1 MVSSCGSVCSDSQCGLG.....PTTCRTCTCFHPICCGSSCC 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCtUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	412	27.8	1417	4	US-08-900-230-3	Sequence 3, Appli	
2	412	27.8	2211	4	US-09-738-884-1	Sequence 1, Appli	
3	391.5	26.4	1345	2	US-08-977-767-3	Sequence 3, Appli	
4	378	25.5	1400	4	US-08-630-915A-37	Sequence 37, Appli	
5	365	24.6	1917	4	US-09-627-650B-5	Sequence 5, Appli	
6	365	24.6	1917	4	US-09-436-063C-5	Sequence 5, Appli	
7	338.5	22.8	2088	4	US-09-436-063C-5	Sequence 13, Appli	
8	338.5	22.8	2088	4	US-09-548-372D-13	Sequence 13, Appli	
9	316.5	21.3	1128	4	US-09-548-367D-13	Sequence 13, Appli	
10	316.5	21.3	1128	4	US-09-627-650B-11	Sequence 11, Appli	
11	316.5	21.3	1652	4	US-09-436-063C-11	Sequence 11, Appli	
12	316.5	21.3	1652	4	US-09-627-650B-11	Sequence 1, Appli	
13	316.5	21.3	2508	4	US-09-436-063C-1	Sequence 1, Appli	
14	316.5	21.3	2508	4	US-09-627-650B-7	Sequence 7, Appli	
15	316.5	21.3	2544	4	US-09-436-063C-7	Sequence 7, Appli	
16	316.5	21.3	2544	4	US-09-627-650B-3	Sequence 3, Appli	
17	316.5	21.3	2601	4	US-09-436-063C-3	Sequence 3, Appli	
18	316.5	21.3	2601	4	US-09-627-650B-9	Sequence 9, Appli	
19	282	19.0	801	1	US-07-906-349A-6	Sequence 9, Appli	
20	209.5	14.1	430	4	US-07-936-349A-6	Sequence 6, Appli	
21	204	13.7	430	4	US-09-156-835B-2	Sequence 2, Appli	
22	198	13.3	1104	3	US-08-997-897-2	Sequence 2, Appli	
23	198	13.3	1104	2	US-08-327-837-5	Sequence 5, Appli	
24	184.5	12.4	120	3	US-08-828-584-5	Sequence 5, Appli	
25	180.5	12.2	2556	1	US-08-508-761B-22	Sequence 22, Appli	
26	180.5	12.2	2556	1	US-08-083-590A-20	Sequence 20, Appli	
27	179.5	12.1	1443	4	US-08-532-384-20	Sequence 20, Appli	
					US-08-990-823-112	Sequence 112, App	

28	176.5	11.9	2556	1	US-08-185-432-17	Sequence 17, Appl
29	176.5	11.9	2556	4	US-08-899-232-2	Sequence 2, Appl
30	169	11.4	578	4	US-08-981-392-13	Sequence 13, Appl
31	169	11.4	1964	4	US-09-467-997-1	Sequence 1, Appl
32	169	11.4	2050	2	US-08-347-594A-2	Sequence 2, Appl
33	162	10.9	109	2	US-08-527-044-2	Sequence 2, Appl
34	162	10.9	109	3	US-09-013-780-2	Sequence 2, Appl
35	162	10.9	341	2	US-07-209-521-11	Sequence 11, Appl
36	162	10.9	589	1	US-07-668-648-2	Sequence 2, Appl
37	162	10.9	589	2	US-08-429-998-2	Sequence 2, Appl
38	162	10.9	589	2	US-08-431-333-2	Sequence 2, Appl
39	162	10.9	589	5	PCR-US91-02331-2	Sequence 2, Appl
40	162	10.9	1610	5	PCR-US95-11684-4	Sequence 4, Appl
41	160.5	10.8	2199	5	PCR-US95-11684-2	Sequence 2, Appl
42	159	10.7	2471	1	US-08-185-432-16	Sequence 16, Appl
43	159	10.7	2471	1	US-08-083-590A-19	Sequence 19, Appl
44	159	10.7	2471	3	US-08-532-384-19	Sequence 19, Appl
45	159	10.7	2471	4	US-08-899-232-1	Sequence 1, Appl

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

Query Match	27.8%	Score 412;	DB 4;	Length 1417;
Best Local Similarity	33.1%	Pred. No. 8.8e-24;		
Matches	92;	Conservative	13;	Mismatches 97;
				Indels 76;
				Gaps 22;

5 CCGSVCSHOSCGUGL-----GOESCCRPSCCO-----TTCCRTT----- 38

Db 899 CCGGGCCTC"ACGGGCCTCTGGGTCCGCACACGGCTCATCTGTGCTTC"GGTACG 958

Qy 39 -CCRPSCCISSCOR--PSCCISSCCKPSCCLTTCRTTCRPPSCCISSCCRPSCCISSCC 95
Db 959 GCC--GCTTCGCTTCACGCGCGC--ACCTACGCTGCC--GCTGCGCTCACACATGCC 1012
Qy 96 KPSGCRRTTCRPPSCCISSCCRPSCCISSCCKPSCCCTTC--CRPSCCISSC-C 145
Db 1013 TGGCC--TAC--GCCAACTCTGCTCAAC--CCGCTCGTCTACGCGCTCGCTCGCGC 1065
Qy 146 RPSCCISSCCKPSC--CQTTC--RPSCCISSCYRPOCCOPSCCRP-ACC-IS 192
Db 1066 ---CACTTCGCGCGCGC--TTCCGCGCGCTGTGCGGTGC--GGCC--GCCAGCGCGC 1117
Qy 193 SCCHPSCVSCRCFPSCPTTCRTTCFHPITCCGSSCC 230
Db 1118 ACCGTGCGCGCGC--GCCITGC--GTCGCTCCGCGCC 1152

RESULT 2

US-09-738-884-1
; Sequence 1, Application US/09738884
; Patent No. 6391606
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000849
; CURRENT APPLICATION NUMBER: US/09/738,884
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2211
; TYPE: PRT
; ORGANISM: Human
US-09-738-884-1

Query Match 27.8%; Score 412; DB 4; Length 2211;

Best Local Similarity 29.9%; Pred. No. 1.2e-23;
Matches 88; Conservative 19; Mismatches 99; Indels 88; Gaps 19;

Qy 3 SSCGVS-----CSDSC-----GGLGQESCCR-----PSCQTTTC--CRTTCR 41
Db 973 TCGCGCTCGTGAGCTGAGCTGCTGGAGGGCCAGGAGGAGCGCTCATATCAT 1032
Qy 42 PSCCISSCCRPSCCISSCCKPSC--LTTTCRTTCCHPS-----CCISSCCRPSCC 91
Db 1033 GGCATAC-----CCTCAC--TCCAAGATTCTCTCGGAGCGTGGCCAAAGCGTGGC 1086
Qy 92 SSCCKPSCRTTCRPPSCCISSCCRPSCCISSCCKPSCRTTCRPPSCCISSCCRPSCC 151
Db 1087 GACCATGCTTCACG---CTGTCC---CCTTACCTGTGC-ATCTATCCCTGGAGAACCA 1139
Qy 152 SSCCKPSC-----COTTCCRPSCCISSCYRPOCCOPSCCRPACCISSC----- 194
Db 1140 CTGGGGCTGGAGCAGAGCTGCGATGGCCCGCGC---ACC--TCTGCACCATCTGGGG 1194
Qy 195 --CHPSC-----CVSSRCRPFSCPTTC--RTTCFHPITCCG-----SSCC 230
Db 1195 GACATGCTGGTGACACAGCGCTGACCTCCCAATC-----CCGAGGAGCTGCC 1244

RESULT 3

US-08-977-767-3
; Sequence 3, Application US/08977767
; Patent No. 5972684
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Greenwald, Sara
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII

; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,767
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0423 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1345 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1532042
US-08-977-767-3

Query Match 26.4%; Score 391.5; DB 2; Length 1345;

Best Local Similarity 30.3%; Pred. No. 2.8e-22;
Matches 86; Conservative 22; Mismatches 105; Indels 71; Gaps 19;

Qy 5 CCG-----SVCSDSGCGGLGQESCCRPPSCCOTTCRTTCRPPSCC1-SSCCRP-----SCC 55
Db 858 CCGCATCTCTTAAGAATGATGCC--TACTTTC--TTCAAGACCTGAGCTGAGATGCC 913
Qy 56 ISS---GC-KPSCGLTT-----CCTTCRPPSC-CISSCCRPSCCISSCCKPSC- 99
Db 914 TGTGCGCGAGTCTTTGGCTTCATCATCAGGGCTCTCTCAGACACCTGAGTCT 973
Qy 100 -----CRTT--CCRPSCCISSCCRPSCCISSCCKPSC-----C 130
Db 974 CGGAGACTGTATCCTGGATCTCATGACAGGGCCCTCAATATCACCTCCCTCAGTGC 1033
Qy 131 RTTCRPPSC-CISSCCRPSC--ISSCKPSC--QTTCRPPSCCISSCYRPPSCCKR 186
Db 1034 ATTCCTCCGAGCTTCTGAGCCAGATCC--TCCGGTCCCAGTCTTCAGAGCTCAGTCCG 1091
Qy 187 PACCISSCCRPSCVSSRCRPFSCPTTCRTTCFHPITCCGSSCC 230
Db 1092 TAAC-GGCCGCGCTGCA---GC---CCTTCGCGCAGCGGCC 1128

RESULT 4

US-08-630-915A-37
; Sequence 37, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, NO. 6309820H
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.

APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-37

Query Match 25.5%; Score 378; DB 4; Length 1400;
Best Local Similarity 28.8%; Pred. No. 2.9e-21;
Matches 88; Conservative 23; Mismatches 95; Indels 100; Gaps 22;

QY 5 CGGVCDSQSGGGLGQSCCR-----PSC--CQTTCCRTTCRPSCCGSSCCR 51
DB 369 CGCTC-----CA-GTGAACCACTGACTGATTCACATCTGCCCTGCC---CCCAACT 420
QY 52 PSCCISS-----CCKPSCCLTT-----CCR--TTCCRPSCCISCCRPSCC-- 90
DB 421 GSCCTTGCGTGAGACCCCGCCCTTGSCAGTAACCTTTCAGAGCCCTCCACGACCC 480
QY 91 ----ISSCKPSCCR--TTCCRPSCCISCCRPSCCISCCRPSCRTT-----CCR--- 136
DB 481 TAATAACTGGGCGGACTTCAGTCC--ACGTGGCCACGACGACCAATGAGAAACGAA 538
QY 137 -----PSCCISCCRPSCCISCCRPSCC--TTCCRPSCCI 171
DB 539 ACGGATACTGGGATGCGATGGCAGCCAGCCCTCTCACCCTTCCCAAGTCC--GGCA 597
QY 172 SSCYRPPCCQ--SCCRPACCISSCHPSCCVSSCCPFPSCP-----TTCCRTTFCFPICC 225
DB 598 GTTAAGGAGAGGTCC--CCCTTAC--TCCAGCCACGGCCACTGGCTCC---TC---CCC 648
QY 226 GS-SCC 230
DB 649 GTCTCC 654

RESULT 5
US-09-627-650B-5
Sequence 5, Application US/09627650B
Patent No. 6406872
GENERAL INFORMATION:
APPLICANT: Bamber, Bruce

APPLICANT: Jorgensen, Erik
TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
TITLE OF INVENTION: Methods Related Thereto
FILE REFERENCE: 21101.000903
CURRENT APPLICATION NUMBER: US/09/627,650B
CURRENT FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 09/436,063
PRIOR FILING DATE: 1999-11-08
PRIOR APPLICATION NUMBER: 60/107,727
PRIOR FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1917
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-627-650B-5
Query Match 24.6%; Score 365; DB 4; Length 1917;
Best Local Similarity 27.1%; Pred. No. 3.4e-20;
Matches 84; Conservative 34; Mismatches 92; Indels 100; Gaps 20;
QY 5 CCG-----SVCSID--OSCGQLGQSCCRPSCCQTTCC-----RPTTCRPSCC 45
DB 1281 CCGTCGGATATCAGCAACGGATGAAGCTTGCCGAGCCAGAAAGAAATCTCGAATGC 1340
QY 46 ISSCKRPSCCISCKPSCCLT--TCRTTCCRPSCCISCCRPSC-CISSCKPSC- 100
DB 1341 TGACC--CCTTTAGCATCTTGACTCTCTTCC--TCCAAACGTACTCTATCGCTTCC 1396
QY 101 -----RPTTCRPSCCISCCRPSCCISCKPSCRTT---CCRPSCCISCCRPSC- 149
DB 1397 TCGTATTTC--ACAACACC---ACGTACC--GCCCGTTTACTCGTCC--ACGATCAAA 1448
QY 150 ---CISSCKPSCCQTTCCRPSC-----CCISSCYRPQ-----CCQ 181
DB 1449 CGTCAACCTGTACATTCGGAGTCGAGCGCAGCAGCAGATTTCTCAATGAGATGCGAG 1508
QY 182 PSCC---RPACCISSCHPSCCVS-----SCRPFSCTTTC--CRPTTC 219
DB 1509 TCCGAATGAACCTAAC---TCCAAATGCTCGGACGGAGTAACCTCACAAGCATCTATTTTC 1565
QY 220 FPITCGSSSC 229
DB 1566 TGTATCAGAC 1575
RESULT 6
US-09-436-063C-5
Sequence 5, Application US/09436063C
Patent No. 6407210
GENERAL INFORMATION:
APPLICANT: Bamber, Bruce
APPLICANT: Jorgensen, Erik
TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
TITLE OF INVENTION: Methods Related Thereto
FILE REFERENCE: P-1095corrected
CURRENT APPLICATION NUMBER: US/09/436,063C
CURRENT FILING DATE: 1999-11-08
PRIOR APPLICATION NUMBER: 60/107727
PRIOR FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1917
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-436-063C-5
Query Match 24.6%; Score 365; DB 4; Length 1917;
Best Local Similarity 27.1%; Pred. No. 3.4e-20;
Matches 84; Conservative 34; Mismatches 92; Indels 100; Gaps 20;


```

: FILE REFERENCE: 21101.000903
: CURRENT APPLICATION NUMBER: US/09/627,650B
: CURRENT FILING DATE: 2000-07-28
: PRIOR APPLICATION NUMBER: 09/436,063
: PRIOR FILING DATE: 1999-11-08
: PRIOR APPLICATION NUMBER: 60/107,727
: PRIOR FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1128
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
: US-09-627-650B-11

Query Match          21.3%; Score 316.5; DB 4; Length 1128;
Best Local Similarity 25.1%; Pred. No. 9.2e-17;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;

QY 4 SCGGSVCSDQSGGGLG---QESCCRPSCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 721 TCCG-CCTTGCCGAATGCACATCAACGCGGTCCAAACCTCGATCATCGCGGTCAATCAAG 779

QY 60 CKPSCCLTTCRPTCCRPSCCISSC-CRPSCCISS-----CCKPSCC---RTT 103
Db 780 CACTCGAATCGATTCTGTCTGACTCAAGTCACATTCACATCTGCGTGAATTCGCGCGTTT 839

QY 104 CCR-----PSCCISSC-CRPSCCISSCCRPSCRTTCCRPSCCISSCCRP 147
Db 840 CCTCTTGTTTCATCTTGTTCAACACTCTCTTCTGCGTGAATTC-TACTGTACAAATCCAA 898

QY 148 SC-CISSCC-----KPSCCQTTCCRPSCCISSCYRPOCCOPSCC 185
Db 899 GCGTCTGCCGTATATTAGTCAACACGAGCGTGACCGTTGCGATGC-----TCCAGACC 951

QY 186 RPAC-----CISSCCHPSCCVSSCRCPSC--PTTCCRT-----TCFHPICCG 226
Db 952 TTCATTAATCTCAATCC--AACTTCTCTCA-TCATTTTCCATTTTCGAATATCTCTTTTCT 1008

QY 227 SSC 229
Db 1009 TGC 1011

RESULT 10
US-09-436-063C-11
: Sequence 11, Application US/09436063C
: Patent No. 6407210
: GENERAL INFORMATION:
: APPLICANT: Bamber, Bruce
: APPLICANT: Jorgensen, Erik
: TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
: TITLE OF INVENTION: Methods Related Thereto
: FILE REFERENCE: P-1095corrected
: CURRENT APPLICATION NUMBER: US/09/436,063C
: CURRENT FILING DATE: 1999-11-08
: PRIOR APPLICATION NUMBER: 60/107727
: PRIOR FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1128
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
: US-09-436-063C-11

Query Match          21.3%; Score 316.5; DB 4; Length 1128;
Best Local Similarity 25.1%; Pred. No. 9.2e-17;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;

QY 4 SCGGSVCSDQSGGGLG---QESCCRPSCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 721 TCCG-CCTTGCCGAATGCACATCAACGCGGTCCAAACCTCGATCATCGCGGTCAATCAAG 779

```

QY	60	CKPSCCLTTCCRTTTCCKPSCCISSC-CRPSCCISS-----CKKPSAC---RPT 103
Db	780	CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTGACATCTGCAGCCGTGGCGGTTTT 839
QY	104	CCR-----PSCCISSC-CRPSCCISSCCRPSCCRTTCCRPSCCISSCRIP 147
Db	840	CCCTCTTTTTCATCTTGTTCAACACTCTCTTCTGGCTGATTC-TACTGTACAAATGCCAA 898
QY	148	SC-CIISSC-----KPSCCOTTCCRPSCCISSCYRPOCCUPSCC 185
Db	899	CGCTCTCCCGTATATATAGTCAACACGAGGTGACCGTTGCCATGC-----TCCACACC 951
QY	186	RPAC-----CISSCHPSCVSSCKRPSC--PTTCCRT-TCFHPICCG 226
Db	952	TTCATTAATCTCAATCC--AATCTCCCA-TCATTTTCCATTTCCGAATATCTCTTTTCT 1008
QY	227	SSC 229
Db	1009	TGC 1011
 RESULT 11 US-09-627-650B-1		
:	:	; Sequence 1, Application US/09627650B
:	:	; Patent No. 6406872
:	:	; GENERAL INFORMATION:
:	:	; APPLICANT: Bamber, Bruce
:	:	; APPLICANT: Jorgensen, Erik
:	:	; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
:	:	; TITLE OF INVENTION: Methods Related Thereto
:	:	; FILE REFERENCE: 21101.000903
:	:	; CURRENT APPLICATION NUMBER: US/09/627,650B
:	:	; CURRENT FILING DATE: 2000-07-28
:	:	; PRIOR APPLICATION NUMBER: 09/436,063
:	:	; PRIOR FILING DATE: 1999-11-08
:	:	; PRIOR APPLICATION NUMBER: 60/107,727
:	:	; PRIOR FILING DATE: 1998-11-09
:	:	; NUMBER OF SEQ ID NOS: 50
:	:	; SOFTWARE: PatentIn Ver. 2.1
:	:	; SEQ ID NO 1
:	:	; LENGTH: 1652
:	:	; TYPE: PRT
:	:	; ORGANISM: Caenorhabditis elegans
US-09-627-650B-1		
 Query Match 21.3% Score 316.5; DB 4; Length 1652;		
Best Local Similarity 25.1%; Pred. No. 1.2e-16;		
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;		
QY	4	SCCGSVCSQSUGUGLA--QESCCRPSCGUTTCRRITC-CRPSCCISSCCRPSCCISSC 59
Db	1222	TCCG-CCTTGCGGAATGCACATGCAAGCGGGTCCAACCTCGATCATCGCGGTATTAAG 1280
QY	60	CKPSCCLTTCCRTTTCCKPSCCISSC-CRPSCCISS-----CKKPSAC---RPT 103
Db	1281	CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTGACATCTGCAGCCGTGGCGGTTTT 1340
QY	104	CCR-----PSCCISSC-CRPSCCISSCCRPSCCRTTCCRPSCCISSCRIP 147
Db	1341	CCCTCTTTTTCATCTTGTTCAACACTCTCTTCTGGCTGATTC-TACTGTACAAATGCCAA 1399
QY	148	SC-CIISSC-----KPSCCOTTCCRPSCCISSCYRPOCCUPSCC 185
Db	1400	CGCTCTCCCGTATATATAGTCAACACGAGGTGACCGTTGCCATGC-----TCCACACC 1452
QY	186	RPAC-----CISSCHPSCCVSSCKRPSC--PTTCCRT-TCFHPICCG 226
Db	1453	TTCATTAATCTCAATCC--AATCTCCCA-TCATTTTCCATTTCCGAATATCTCTTTTCT 1509
QY	227	SSC 229
Db	1510	TGC 1512

```
RESULT 12
US-09-436-063C-1
; Sequence 1, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1652
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-1

Query Match      21.3%  Score 316.5;  DB 4;  Length 1652;
Best Local Similarity 25.1%  Pred. No. 1.2e-16;
Matches 76;  Conservative 27;  Mismatches 111;  Indels 89;  Gaps 17;

QY 4  SCGSGVCSQDSQGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 1222 TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGGTCAAG 1280

QY 60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db 1281 CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGGTGCGCGTTT 1340

QY 104  CCR-----PSCCISSC-CRPSCCISSCCRPSCCRTTCRPSCCISSCCRP 147
Db 1341 CCTCTTGTTCATCTGTGTCAGTCACAGTCACATCTCTCTGGCTGATTC-TACTGTACAATCAA 1399

QY 148  SC-CISSC-----KPSCCOTTCCRPSCCISSCVRPQCCQPSCC 185
Db 1400 CGCTGCGCGTATATTAGTAACACGAGGGTGACCGTTGCGATGC-----TCCAGACC 1452

QY 186  RPAC-----CISSCCHPSCCVSSCRCPSC--PTTCCRT-----TCFHPICCG 226
Db 1453 TTCATTAATCTCAATCC--AACTTCTCA-TCAATTTTCCATTTCCGAATATCTCTTTTCT 1509

QY 227  SSC 229
Db 1510 TGC 1512

RESULT 13
US-09-627-650B-7
; Sequence 7, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.000903
; CURRENT APPLICATION NUMBER: US/09/627.650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508

QY 4  SCGSGVCSQDSQGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 2106 TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGGTCAAG 2164

QY 60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db 2165 CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGGTGCGCGTTT 2224

QY 104  CCR-----PSCCISSC-CRPSCCISSCCRPSCCRTTCRPSCCISSCCRP 147
Db 2225 CCTCTTGTTCATCTGTGTCAGTCACAGTCACATCTCTCTGGCTGATTC-TACTGTACAATCAA 2283

QY 148  SC-CISSC-----KPSCCOTTCCRPSCCISSCVRPQCCQPSCC 185
Db 2336 TCCG-CCTTGGCGAATGCACATGCAACGGCGGTGACCGTTGCGATGC-----TCCAGACC 2393

QY 186  RPAC-----CISSCCHPSCCVSSCRCPSC--PTTCCRT-----TCFHPICCG 226
Db 2337 TTCATTAATCTCAATCC--AACTTCTCA-TCAATTTTCCATTTCCGAATATCTCTTTTCT 2393

QY 227  SSC 229
Db 2394 TGC 2396

RESULT 14
US-09-436-063C-7
; Sequence 7, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-7

Query Match      21.3%  Score 316.5;  DB 4;  Length 2508;
Best Local Similarity 25.1%  Pred. No. 1.6e-16;
Matches 76;  Conservative 27;  Mismatches 111;  Indels 89;  Gaps 17;

QY 4  SCGSGVCSQDSQGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 2106 TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGGTCAAG 2164

QY 60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db 2165 CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGGTGCGCGTTT 2224

QY 104  CCR-----PSCCISSC-CRPSCCISSCCRPSCCRTTCRPSCCISSCCRP 147
Db 2225 CCTCTTGTTCATCTGTGTCAGTCACAGTCACATCTCTCTGGCTGATTC-TACTGTACAATCAA 2283

QY 148  SC-CISSC-----KPSCCOTTCCRPSCCISSCVRPQCCQPSCC 185
Db 2336 TCCG-CCTTGGCGAATGCACATGCAACGGCGGTGACCGTTGCGATGC-----TCCAGACC 2393

QY 186  RPAC-----CISSCCHPSCCVSSCRCPSC--PTTCCRT-----TCFHPICCG 226
Db 2337 TTCATTAATCTCAATCC--AACTTCTCA-TCAATTTTCCATTTCCGAATATCTCTTTTCT 2393

QY 227  SSC 229
Db 2394 TGC 2396
```

```
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-7

Query Match      21.3%  Score 316.5;  DB 4;  Length 2508;
Best Local Similarity 25.1%  Pred. No. 1.6e-16;
Matches 76;  Conservative 27;  Mismatches 111;  Indels 89;  Gaps 17;

QY 4  SCGSGVCSQDSQGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 2106 TCCG-CCTTGGCGAATGCACATGCAACGGCGGTGACCGTTGCGATGC-----TCCAGACC 2164

QY 60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db 2165 CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGGTGCGCGTTT 2224

QY 104  CCR-----PSCCISSC-CRPSCCISSCCRPSCCRTTCRPSCCISSCCRP 147
Db 2225 CCTCTTGTTCATCTGTGTCAGTCACAGTCACATCTCTCTGGCTGATTC-TACTGTACAATCAA 2283

QY 148  SC-CISSC-----KPSCCOTTCCRPSCCISSCVRPQCCQPSCC 185
Db 2284 CGCTGCGCGTATATTAGTAACACGAGGGTGACCGTTGCGATGC-----TCCAGACC 2336

QY 186  RPAC-----CISSCCHPSCCVSSCRCPSC--PTTCCRT-----TCFHPICCG 226
Db 2337 TTCATTAATCTCAATCC--AACTTCTCA-TCAATTTTCCATTTCCGAATATCTCTTTTCT 2393

QY 227  SSC 229
Db 2394 TGC 2396

RESULT 14
US-09-436-063C-7
; Sequence 7, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-7

Query Match      21.3%  Score 316.5;  DB 4;  Length 2508;
Best Local Similarity 25.1%  Pred. No. 1.6e-16;
Matches 76;  Conservative 27;  Mismatches 111;  Indels 89;  Gaps 17;

QY 4  SCGSGVCSQDSQGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 2106 TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGGTCAAG 2164

QY 60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db 2165 CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGGTGCGCGTTT 2224

QY 104  CCR-----PSCCISSC-CRPSCCISSCCRPSCCRTTCRPSCCISSCCRP 147
Db 2225 CCTCTTGTTCATCTGTGTCAGTCACAGTCACATCTCTCTGGCTGATTC-TACTGTACAATCAA 2283

QY 148  SC-CISSC-----KPSCCOTTCCRPSCCISSCVRPQCCQPSCC 185
Db 2336 TCCG-CCTTGGCGAATGCACATGCAACGGCGGTGACCGTTGCGATGC-----TCCAGACC 2393

QY 186  RPAC-----CISSCCHPSCCVSSCRCPSC--PTTCCRT-----TCFHPICCG 226
Db 2337 TTCATTAATCTCAATCC--AACTTCTCA-TCAATTTTCCATTTCCGAATATCTCTTTTCT 2393

QY 227  SSC 229
Db 2394 TGC 2396
```

1b 2284 GCGTCTGCCGTATATTAGTGAACACAGAGGGTGACCGTTGCGATGC-----TCCAGACC 2336
QY 186 RPAC-----CISGCHPSCVSSCRCPFSC--PTTCRT-----TCFHPICCG 226
Db 2337 TTCATTAAATCTCAATCC--AACTTCTCA-TCAATTTCCATTTCGAATATCTCTTTTCT 2393
QY 227 SSC 229
Db 2394 TGC 2396

RESULT 15

US-09-627-650B-3
: Sequence 3, Application US/09627650B
: Patent No. 6406872
: GENERAL INFORMATION:
: APPLICANT: Bamber, Bruce
: APPLICANT: Jorgensen, Erik
: TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
: TITLE OF INVENTION: Methods Related Thereto
: FILE REFERENCE: 21101.000903
: CURRENT APPLICATION NUMBER: US/09/627,650B
: PRIOR FILING DATE: 2000-07-28
: PRIOR APPLICATION NUMBER: 09/436,063
: PRIOR FILING DATE: 1999-11-08
: PRIOR APPLICATION NUMBER: 60/107,727
: PRIOR FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 2544
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
US-09-627-650B-3

Query Match 21.3%; Score 316.5; DB 4; Length 2544;
Best Local Similarity 25.1%; Pred. No. 1.6e-16;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;
QY 4 SCGGSVCDSGCGGLG---QESCRPSCGTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 2142 TCGG-CGTTGCCGAATGCACATGCAGCGGCTCCAACCTCGATCATCGCCGTCATCAAG 2200
QY 60 CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISS-----CKKPSCC---RTT 103
Db 2201 CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTGACATCGTCAGCGCGCGGTTT 2260
QY 104 CCR-----PSCCISSC-CRPSCCISSCCRPSCCCTTCRPSCCISSCCRP 147
Db 2261 CCTCTTTGTTTCATCTTGTTCACACTCTCTTGGCTGATTC-TACTGTACAAATCCAA 2319
QY 148 SC-CISSC-----KPSCCOTTCRPSCCISSCYRPOCCQPSCC 185
Db 2320 CGGTCTGCCGTATATTAGTGAACACAGGGTGACCGTTGCCATGC-----TCCAGACC 2372
QY 186 RPAC-----CISGCHPSCVSSCRCPFSC--PTTCRT-----TCFHPICCG 226
Db 2373 TTCATTAAATCTCAATCC--AACTTCTCA-TCAATTTCCATTTCGAATATCTCTTTTCT 2429
QY 227 SSC 229
Db 2430 TGC 2432

Search completed: June 11, 2003, 11:06:58
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 11:02:46 : Search time 78 Seconds
(without alignments)
607.575 Million cell updates/sec

Title: US-09-874-062-3
Perfect score: 1484
Sequence: 1 MVSSCGSVCSQSGQGLG.....PTTCRTTCFHPICGSSCC 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp-invertebrate.*
- 6: sp-mammal.*
- 7: sp-mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-virus.*
- 13: sp-vertebrate.*
- 14: sp-unclassified.*
- 15: sp-virus.*
- 16: sp-bacteriaph.*
- 17: sp-archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	996.5	67.1	195	11 Q9D0X9	Q9d0x9 mus musculus
2	968	65.2	210	4 Q9BYR0	Q9byr0 homo sapien
3	925	62.3	201	4 Q9BQ66	Q9bq66 homo sapien
4	898	60.5	195	4 Q9BYQ6	Q9byq6 homo sapien
5	866.5	58.4	191	4 Q9BYQ8	Q9byq8 homo sapien
6	858.5	57.9	193	4 Q9BYQ5	Q9byq5 homo sapien
7	808.5	54.5	202	11 Q91W93	Q91w93 mus musculus
8	806	54.3	165	11 Q9D7P3	Q9d7p3 mus musculus
9	804.5	54.2	186	4 Q9BYR2	Q9byr2 homo sapien
10	800.5	53.9	166	4 Q9BYR3	Q9byr3 homo sapien
11	764.5	51.5	168	11 Q9D732	Q9d732 mus musculus
12	744	50.1	195	11 Q9D141	Q9d141 mus musculus
13	738	49.7	167	11 Q9D122	Q9d122 mus musculus
14	709	47.8	191	11 Q9D3H7	Q9d3h7 mus musculus
15	701.5	47.3	186	11 Q64526	Q64526 mus musculus
16	693.5	46.7	223	11 Q62220	Q62220 mus musculus

17	684	46.1	230	11 Q64507	Q64507 mus musculus
18	654.5	44.1	136	4 Q9BYR5	Q9byr5 homo sapien
19	645	43.5	98	4 Q9BYR4	Q9byr4 homo sapien
20	619	41.7	151	6 Q28576	Q28576 ovis aries
21	584.5	39.4	127	4 Q9BYQ7	Q9byq7 homo sapien
22	584.5	39.4	174	4 Q9BYQ4	Q9byq4 homo sapien
23	569.5	38.4	126	6 Q28707	Q28707 oryctolagus
24	560	37.7	188	11 Q70148	Q70148 rattus norv
25	546	36.8	227	11 Q9D225	Q9d225 mus musculus
26	545.5	36.8	159	4 Q9BYQ0	Q9byq0 homo sapien
27	542.5	36.6	159	4 Q9BYQ3	Q9byq3 homo sapien
28	530	35.7	154	4 Q9BYQ2	Q9byq2 homo sapien
29	521	35.1	114	4 Q9BYQ9	Q9byq9 homo sapien
30	515	34.7	154	4 Q9BYP9	Q9byp9 homo sapien
31	481.5	32.4	106	4 Q9BYU7	Q9byu7 homo sapien
32	463.5	31.2	175	4 Q07628	Q07628 homo sapien
33	451	30.4	177	4 Q07627	Q07627 homo sapien
34	447	30.1	174	4 Q9BYS1	Q9bys1 homo sapien
35	438.5	29.5	167	4 Q9BYS2	Q9bys2 homo sapien
36	432	29.1	182	6 Q29619	Q29619 ovis aries
37	431.5	29.1	169	4 Q14564	Q14564 homo sapien
38	428.5	28.9	189	11 Q9D527	Q9d527 mus musculus
39	421.5	28.4	177	11 Q9D644	Q9d644 mus musculus
40	380	25.6	130	11 Q9Z287	Q9z287 mus musculus
41	376.5	25.4	111	4 Q9BYQ1	Q9byq1 homo sapien
42	375	25.3	661	5 Q17982	Q17982 caenorhabdi
43	366.5	24.7	136	11 Q9D718	Q9d718 mus musculus
44	360.5	24.3	152	6 Q29620	Q29620 ovis aries
45	358	24.1	251	5 Q24774	Q24774 enclytraeus

ALIGNMENTS

RESULT 1

Q9D0X9	PRELIMINARY;	PRT: 195 AA.
ID Q9D0X9		
AC Q9D0X9		
DT 01-JUN-2001 (TREMBLrel. 17, Created)		
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE I110054P19RIK protein.		
GN I110054P19RIK.		
OS Mus musculus (Mouse).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.		
OX NCBI_TaxID=10090;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;		
RX MEDLINE=21085660; PubMed=11217851;		
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,		
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,		
RA Kuehl P., Lewis S., Matsuo Y., Nikaido T., Pesole G., Quackenbush J.,		
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA Nordone P., Ring B., Kingwald M., Rodriguez I., Sakamoto N.,		
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Willing L.,		
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,		
RA Hayashizaki Y.;		
RT "Functional annotation of a full-length mouse cDNA collection.";		
RL Nature 409:685-690(2001).		
DR EMBL; AK004258; BAB23238.1; ..		
DR HSSP; Q46655; 1CJH.		

seq ID NO:2, Result 2
Database: GenEmbl; AC NO: AC006070

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-595122"
BASE COUNT 107326 A 64859 C 63333 G 108900 T 3154 others
ORIGIN

Query Match      82.6%  Score 909.2  DB 2:  Length 347572;
Best Local Similarity 91.9%  Pred. No. 6e-148;
Matches 1002; Conservative 0; Mismatches 13; Indels 75; Gaps 1;

QY 1 CTGGGAACCCACCCAGAACCTCCACCCCTCTGACGCCATGTCAGCTCTCTGTCGTGCTCT 60
DB 105475 CTGGGAACCCACCCAGAACCTCCACCCCTCTGACGCCATGTCAGCTCTCTGTCGTGCTCT 105534
QY 61 GTCTGCTCTGACAGAGCTGTGTCAGGCTCTGCGCCAGGAGAGCTCTGCGCCGCCAGC 120
DB 105535 GTCTGCTCTGACAGAGCTGTGTCAGGCTCTGCGCCAGGAGAGCTCTGCGCCGCCAGC 105594
QY 121 TGCTGCCAGACCACTGCTGTCAGGACCACTGCTGCGCCGCCAGGCTCTGTCATTTCCAGT 180
DB 105595 TGCTGCCAGACCACTGCTGTCAGGACCACTGCTGCGCCGCCAGGCTCTGTCATTTCCAGT 105654
QY 181 TGCTGCCAGGCTCTGCTGTCATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCAACC 240
DB 105655 CGCTGCCAGGCTCTGCTGTCATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCAACC 105714
QY 241 TGCTGCCAGGACCACTGCTGCGCCGCCAGCTGCTGCAATTTCCAGTCTGTCAGGCTCTCC 300
DB 105715 TGCTGCCAGGACCACTGCTGCGCCGCCAGCTGCTGCAATTTCCAGTCTGTCAGGCTCTCC 105774
QY 301 TGCTGTCATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAAAACCCAGCTGCTGTCGCCGCCAGC 360
DB 105775 TGCTGTCATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAAAACCCAGCTGCTGTCGCCGCCAGC 105818
QY 361 TGCTGTCATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAAAACCCAGCTGCTGTCGCCGCCAGC 420
DB 105819 TGCTGTCATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAAAACCCAGCTGCTGTCGCCGCCAGC 105819
QY 421 TGCTGTCAGGACCACTGCTGCGCCGCCAGCTGCTGTCATTTCTAGTCTGTCAGGCTCTCC 480
DB 105820 TGCTGTCAGGACCACTGCTGCGCCGCCAGCTGCTGTCATTTCTAGTCTGTCAGGCTCTCC 105879
QY 481 TGCTGTCATCTCTAGCTGCTGCAAAACCCAGCTGCTGTCGCCAGACCACTGCTGTCGCCGCCAGC 540
DB 105880 TGCTGTCATCTCTAGCTGCTGCAAAACCCAGCTGCTGTCGCCAGACCACTGCTGTCGCCGCCAGC 105939
QY 541 TGCTGTCATCTCTAGCTGCTGTCAGGACCACTGCTGTCGCCAGGCTCTGTCGCCGCCAGGCT 600
DB 105940 TGCTGTCATCTCTAGCTGCTGTCAGGACCACTGCTGTCGCCAGGCTCTGTCGCCGCCAGGCT 105999
QY 601 TGCTGTCATTTCTAGTCTGTCATCCAGCTGCTGTCAGCTGTCAGCTGTCGCTGCTGCTCTTC 660
DB 106000 TGCTGTCATTTCTAGTCTGTCATCCAGCTGCTGTCAGCTGTCAGCTGTCGCTGCTGCTCTTC 106059
QY 661 AGCTGCCGACCACTGCTGTAGAACCACTGCTGTCACCACTGCTGTCGCGGAGCTCT 720
DB 106060 AGCTGCCGACCACTGCTGTAGAACCACTGCTGTCACCACTGCTGTCGCGGAGCTCT 106119
QY 721 TGCTGTCAGTGAAGCTGCTCTGGAATTTGTCACCTCTTCTGCTCAACCTTCAGTTTCAG 780
DB 106120 TGCTGTCAGTGAAGCTGCTCTGGAATTTGTCACCTCTTCTGCTCAACCTTCAGTTTCAG 106179
QY 781 GCACAGATATCTATTTCAGAACATGTCGACTTCTGCTGATGTCGTCGAAACAGAGGCAATG 840
DB 106180 GCACAGATATCTATTTCAGAACATGTCGACTTCTGCTGATGTCGTCGAAACAGAGGCAATG 106239
QY 841 GACTGATTTGGAATATTTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 106240 GACTGATTTGGAATATTTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 106299
QY 901 TTGAATCTGAATTTACAGTCAAAATCCACATCACATGTTTATGAAATCTTTATTTATTAAT 960
DB 106299 TTGAATCTGAATTTACAGTCAAAATCCACATCACATGTTTATGAAATCTTTATTTATTAAT 106359
```

```
Db 106300 TTGAATCTGAATTTACAGTCAAAATCCACATCACATGTTTATGAAATCTTTATTTATTAAT 106359
QY 961 CAATATACATAAATCTTCAAAATGATATCTCTAGATGTTTCTTCTCTTCTCTTCTCTCTCTCTCT 1020
Db 106360 CAATATACATAAATCTTCAAAATGATATCTCTAGATGTTTCTTCTCTTCTCTCTCTCTCTCTCT 106419
QY 1021 GTATCAATTTTCTATGTCGAAATGTTTCTCTCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1080
Db 106420 GTATCAATTTTCTATGTCGAAATGTTTCTCTCAATCAATCAATCAATCAATCAATCAATCAATCAAT 106479
QY 1081 GCACAAAAAA 1090
Db 106480 GCACAAAAAA 106489

RESULT 2
AC006070
LOCUS Homo sapiens chromosome 17, clone hRPK.206_C_20, linear PRI 17-DEC-1998
DEFINITION AC006070
ACCESSION AC006070.1 GI:4028942
VERSION HTG
KEYWORDS Homo sapiens
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 161987)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.206_C_20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 161987)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J.,
Mychalckey,J., Nahir,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (26-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 161987)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A.,
Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J.,
Mychalckey,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (17-DEC-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Dec 17, 1998 this sequence version replaced gi:4006836.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only the last 161987 base pairs of this clone are being submitted.
```


AC025904
AC025904.2. GI:7382590
HTG: HTGS_PHASE1, HTGS_DRAFT.
KEYWORDS
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-585G19
JOURNAL
Unpublished
REFERENCE
1 (bases 1 to 174032)
2 (bases 1 to 174032)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F., Boguski, L., Bouckgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choelp, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Houghton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larrecque, K., Lamazares, R., Landers, T., Lehoczkv, J., Levine, R., Li, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., Mihov, A., McKernan, K., McPheeters, R., Meidrim, J., Meneses, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zhan, A. and Zody, M.
Direct Submission
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:7249428.
All repeats were identified using RepeatMasker:
Smit, A.F.A., Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
***** Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
***** Project Information
Center project name: L8441
Center clone name: 385_C_19
***** Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.940731
Consensus quality: 158771 bases at least Q40
Consensus quality: 167389 bases at least Q30
Consensus quality: 170660 bases at least Q20
Insert size: 170000; agarose-ff
Insert size: 172432; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-ff
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence report is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2588: contig of 2588 bp in length
* 2589 2688: gap of 100 bp
* 2689 6465: contig of 3777 bp in length
* 6466 6565: gap of 100 bp
* 6566 8846: contig of 2281 bp in length

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 11:02:46 ; Search time 18 Seconds
(without alignments)
1228.384 Million cell updates/sec

Title: US-09-874-062-3

Perfect score: 1484

Sequence: 1 MVSSCGSVCSQSGGGLG.....PTTCRTTCFHPICGSGSCC 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	701.5	47.3	186	2 A45910	ultra-high-sulfur
2	693.5	46.7	223	2 B38346	ultra-high-sulfur
3	684	46.1	230	2 A38346	ultra-high-sulfur
4	619	41.7	151	2 S60314	hair keratin cyste
5	569.5	38.4	126	2 I46489	cysteine-rich hair
6	560	37.7	188	2 JC6547	high-sulfur protei
7	463.5	31.2	175	2 S37649	high-sulfur kerati
8	451	30.4	177	2 S37650	high-sulfur kerati
9	434	29.2	182	1 KRSHHD	keratin high-sulfu
10	432	29.1	182	2 I47105	high-sulfur wool m
11	431.5	29.1	169	1 S18946	ultra high-sulfur
12	408.5	27.5	172	1 KRSHHA	keratin high-sulfu
13	402	27.1	172	2 I47106	high-sulfur wool m
14	384.5	25.9	162	2 I47107	high-sulfur wool m
15	375	25.3	654	2 T30136	hypothetical prote
16	360.5	24.3	152	2 I47108	keratin high-sulfu
17	358.5	24.2	152	1 KRSHHC	keratin high-sulfu
18	358.5	24.2	152	2 I47111	high-sulfur wool m
19	358	24.1	251	2 A55035	cysteine-rich prot
20	350.5	23.6	152	2 I47109	high-sulfur wool m
21	350.5	23.6	152	2 I47112	high-sulfur wool m
22	348	23.5	156	1 KRSHHB	keratin high-sulfu
23	344	23.2	191	2 I46412	keratin high-sulfu
24	325	21.9	197	2 I46413	keratin KAP5.4 - s
25	305.5	20.6	182	2 A36686	ultra-high-sulfur
26	298.5	20.1	572	2 T29880	hypothetical prote
27	296.5	20.0	131	1 KRGT3M	keratin high-sulfu
28	294	19.8	728	2 T20561	hypothetical prote
29	293	19.7	425	2 T18592	hypothetical prote

ALIGNMENTS

RESULT 1

A45910

ultra-high-sulfur keratin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 24-Sep-1999

C:Accession: A45910

R:McNab, A.R.; Wood, L.; Theriault, N.; Gierman, T.; Vogeli, G.

J. Invest. Dermatol. 92, 263-266, 1989

A:Title: An ultra-high sulfur keratin gene is expressed specifically during hair growth

A:Reference number: A45910; MUID:89140394; PMID:2465353

A:Accession: A45910

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-186 <MCN>

A:Cross-references: GB:M27685; MID:g341749; PIDN:AAA81560.1; PID:g1066818

C:Superfamily: ultra-high-sulfur keratin

Query Match 47.3%; Score 701.5; DB 2; Length 186;

Best Local Similarity 49.0%; Pred. No. 1.le-42;

Matches 119; Conservative 20; Mismatches 25; Indels 75; Gaps 11;

QY	1	MVSSCGSVCSQSGGGLGQESCRPSCCOTTCRTTCRRPSC---CISSCCRPSCCIS 57
DB	1	MTNSEC-----SPCQPTCCRTTCRTTCRRPSCVTSQCCQPCSCCGS 45
QY	58	SCCRPSCCLTTCRTTCRRPSCCISSCCRPSCCISSCCRPSCCISSCCRP 117
DB	46	SCCQP-CCQTTCCRT-----CFQPCCVSSCCRTPCCPCCCVSSCCOP 87
QY	118	SCCISSCCRPSCCRTTCRRPSCCISSCCRPSCCISSCCRPSCCISSCCRP 177
DB	88	-----CCQPCCVSSCCRTTCRRPSCCISSCCRPSCCISSCCRPSCCIS 135
QY	178	QCCQPCCRPACCCISSCCRPSCCISSCCRPSCCISSCCRPSCCISSCCRP 227
DB	136	CCQP-----FCLNCCQPCAC-----SGPVTCTR-TCYQTCVCVPCCLSQGCGS 178
QY	228	SCC 230
DB	179	SCC 181

RESULT 2

B38346

ultra-high-sulfur keratin 2 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999

C:Accession: B38660; B38346

R:Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.

J. Biol. Chem. 266, 4024, 1991

A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin

A:Reference number: A38660; MUID:91154184; PMID:1840598